PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07K 14/00 — A2

(11) International Publication Number: WO 98/57984

(43) International Publication Date: 23 December 1998 (23.12.98)

(21) International Application Number: PCT/US98/12776

(22) International Filing Date: 19 June 1998 (19.06.98)

(30) Priority Data:
60/050,244 — 19 June 1997 (19.06.97) — US
09/099,454 — 18 June 1998 (18.06.98) — US

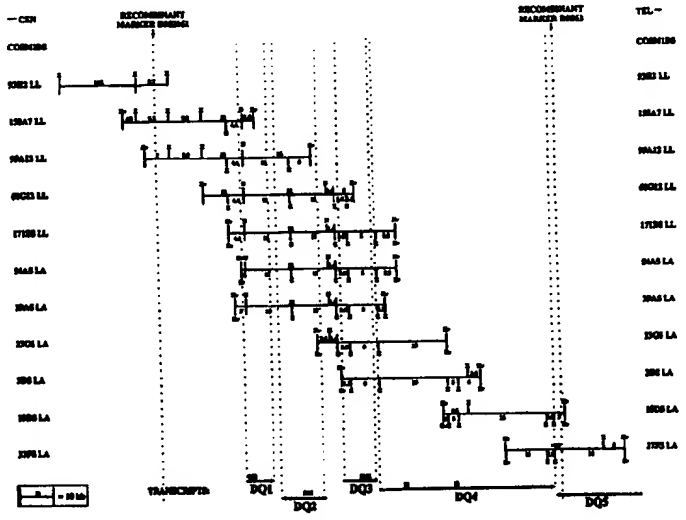
(71)(72) Applicants and Inventors: OZELIUS, Laurie, J. [US/US]; Massachusetts General Hospital–East, Molecular Neurogenetics Unit, Building 149, 13h Street, Charlestow, MA 02129 (US). BREAKEFIELD, Xandra [US/US]; 127 Homer Street, Newton, MA 02159 (US).

(74) Agents: GOLDSTEIN, Jorge, A. et al.; Sterne, Kessler, Goldstein & Fox P.L.L.C., Suite 600, 1100 New York Avenue, N.W., Washington, DC 20005–3934 (US).

(81) Designated States: CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.
With an indication in relation to deposited biological material furnished under Rule 13bis separately from the description.

(54) Title: TORSIN, TORSIN GENES, AND METHODS OF USE



(57) Abstract

The present invention relates, in general, to torsin genes, preferably, torsin A which encodes the torsion dystonia gene, DYT1. In particular, the present invention relates to nucleic acid molecules coding for the torsin protein; purified torsin proteins and polypeptides; recombinant nucleic acid molecules; cells containing the recombinant nucleic acid molecules; antibodies having binding affinity specifically to torsin proteins and polypeptides; hybridomas containing the antibodies; nucleic acid probes for the detection of nucleic acids encoding torsin proteins; a method of detecting nucleic acids encoding torsin proteins or polypeptides in a sample; kits containing nucleic acid probes or antibodies; bioassays using the nucleic acid sequence, protein or antibodies of this invention to diagnose, assess, or prognose a mammal afflicted with torsion dystonia; therapeutic uses; and methods of preventing torsion dystonia in an animal.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

Torsin, Torsin Genes, and Methods of Use

Statement as to Rights to Inventions Made Under Federally-Sponsored Research and Development

Part of the work performed during development of this invention utilized U.S. Government funds. The U.S. Government has certain rights in this invention.

Background of the Invention

Field of the Invention

The present invention relates, in general, to torsin genes, preferably, torsinA which encodes the torsion dystonia gene, DYT1. In particular, the present invention relates to nucleic acid molecules coding for the torsin protein; purified torsin proteins and polypeptides; recombinant nucleic acid molecules; cells containing the recombinant nucleic acid molecules; antibodies having binding affinity specifically to torsin proteins and polypeptides; hybridomas containing the antibodies; nucleic acid probes for the detection of nucleic acids encoding torsin proteins; a method of detecting nucleic acids encoding torsin proteins or polypeptides in a sample; kits containing nucleic acid probes or antibodies; bioassays using the nucleic acid sequence, protein or antibodies of this invention to diagnose, assess, or prognose a mammal afflicted with torsion dystonia; therapeutic uses; and methods of preventing torsion dystonia in an animal (preferably, a human).

Related Art

Movement disorders constitute a group of human neurologic diseases in which aberrant neurotransmission in the basal ganglia is associated with uncontrollable body movements, such as chorea in Huntington disease, tremor and rigidity in Parkinson disease, and twisting contractions in torsion dystonia. Dystonic symptoms can be secondary to a number of neurologic conditions, and to drug or traumatic injury to the brain, but primary or torsion dystonia is distinguished by lack of other neurologic involvement (Fahn, S., *Adv Neurol 50*:1-8 (1988); Chutorian, A. H., *Acta Neuropediatricia 2*:33-45 (1996)) and, in contrast to these other two neurodegenerative diseases, the absence of any distinct neuropathology. The clinical manifestations of dystonia show wide variations in age and site of onset, as well as body regions involved. The prevalence of all forms of primary dystonia is estimated at 3/10,000 in North America (Nutt, J. G., *et al., Mov Disord 3:*188-194 (1988)).

Early onset, generalized dystonia is the most disabling form of primary dystonia. Symptoms usually begin in an arm or leg at around 12 yrs (range 4-44 years) and spread to involve other limbs within about 5 years (Bressman, S. B., *et al., Annal Neurol 36*:771-777 (1994b); Greene, P., *et al., Mov Disord 10:*143-152 (1995)). The clinical spectrum of early onset dystonia is similar in all ethnic populations, with highest prevalence in the Ashkenazi Jewish (termed here AJ) population (Zeman, W., & Dyken, P., *Psychiatr Neurol Neurochir 10:*77-121 (1967); Korczyn, A. D., *et al., Ann Neurol 8:*387-391 (1980); Eldridge, R., *Neurology 20*:1-78 (1970)), due to a founder mutation (Ozelius, L., *et al., Am. J. Hum. Genet. 50:*619-628 (1992); Risch, N. J., *et al., Am J Hum Genet 46:*533-538 (1990) and Risch, N., *et al., Nature Genetics 9:*152-159 (1995)). Early onset dystonia follows an autosomal dominant mode of inheritance with 30-40% penetrance (Bressman, S. B., *et al., Ann Neurol 26*:612-620 (1989); Risch, N. J., *et al., Am J Hum Genet 46:*533-538 (1990)). The responsible gene in Jewish and non-Jewish families has been mapped to human chromosome 9q34 (Ozelius, L., *et al., Neuron 2:*1427-1434 (1989);

Kramer, P. L., *et al., Ann Neurol 27:*114-120 (1990) and Kramer, P., *et al., Am J Hum Gen 55:*468-475 (1994)). Haplotype analysis of the founder mutation in AJ families placed the DYT1 gene in a 1-2 cM interval centromeric to the ASS locus on chromosome 9 (Ozelius, L., *et al., Am. J. Hum. Genet. 50:*619-628 (1992)) with highest lod scores obtained with adjacent markers, D9S62a/b and D9S63 (Risch, N., *et al., Nature Genetics 9:*152-159 (1995)).

Summary of the Invention

The invention provides, in general, isolated nucleic acid molecules coding for torsin, preferably, torsinA which encodes the torsion dystonia gene, DYT1.

The invention further provides purified polypeptides comprising amino acid sequences encoding torsin proteins.

The invention also provides nucleic acid probes for the specific detection of the presence of nucleic acids encoding torsin proteins or polypeptides in a sample.

The invention further provides a method of detecting nucleic acid encoding torsin protein in a sample.

The invention also provides a kit for detecting the presence of nucleic acid encoding torsin protein in a sample.

The invention further provides a recombinant nucleic acid molecule comprising, 5′ to 3′, a promoter effective to initiate transcription in a host cell and the above-described isolated nucleic acid molecule.

The invention also provides a recombinant nucleic acid molecule comprising a vector and the above-described isolated nucleic acid molecule.

The invention further provides a recombinant nucleic acid molecule comprising a sequence complimentary to an RNA sequence encoding an amino acid sequence corresponding to the above-described polypeptide.

The invention also provides a cell that contains the above-described recombinant nucleic acid molecule.

The invention further provides a non-human organism that contains the above-described recombinant nucleic acid molecule.

The invention also provides an antibody having binding affinity specifically to a torsin protein or polypeptide.

The invention further provides a method of detecting torsin protein or polypeptide in a sample.

The invention also provides a method of measuring the amount of torsin protein or polypeptide in a sample.

The invention futher provides a method of detecting antibodies having binding affinity specifically to a torsin protein or polypeptide.

The invention further provides a diagnostic kit comprising a first container means containing the above-described antibody, and a second container means containing a conjugate comprising a binding partner of the monoclonal antibody and a label.

The invention also provides a hybridoma which produces the above-described monoclonal antibody.

The invention further provides diagnostic methods for human disease, in particular, torsion dystonia. Preferably, a method of diagnosing the presence or predisposition to develop torsion dystonia in a patient is provided herein. More preferably, the method comprises:

a) taking a sample from the patient;

b) evaluating the characteristics of torsinA nucleic acid in the sample, wherein the evaluation comprises detecting the GAGGAG region (SEQ ID NO:5 at nucleotide positions 946-951) in the sample; and

c) diagnosing the presence or predisposition to develop torsion dystonia in a patient wherein the absence of a GAG from the GAGGAG region indicates the presence or predisposition to develop torsion dystonia.

The invention also provides methods for therapeutic uses involving all or part of (1) the nucleic acid sequence encoding torsin protein or (2) torsin protein.

Further objects and advantages of the present invention will be clear from the description that follows.

Definitions

In the description that follows, a number of terms used in recombinant DNA (rDNA) technology are extensively utilized. In order to provide a clear and consistent understanding of the specification and claims, including the scope to be given such terms, the following definitions are provided.

Isolated Nucleic Acid Molecule. An "isolated nucleic acid molecule", as is generally understood and used herein, refers to a polymer of nucleotides, and includes but should not be limited to DNA and RNA. The "isolated" nucleic acid molecule is purified from its natural *in vivo* state.

Recombinant DNA. Any DNA molecule formed by joining DNA segments from different sources and produced using recombinant DNA technology (aka. molecular genetic engineering).

DNA Segment. A DNA segment, as is generally understood and used herein, refers to a molecule comprising a linear stretch of nucleotides wherein the nucleotides are present in a sequence that can encode, through the genetic code, a molecule comprising a linear sequence of amino acid residues that is referred to as a protein, a protein fragment or a polypeptide.

Gene. A DNA sequence related to a single polypeptide chain or protein, and as used herein includes the 5′ and 3′ untranslated ends. The polypeptide can be encoded by a full-length sequence or any portion of the coding sequence, so long as the functional activity of the protein is retained.

Complementary DNA (cDNA). Recombinant nucleic acid molecules synthesized by reverse transcription of messenger RNA ("mRNA").

Structural Gene. A DNA sequence that is transcribed into mRNA that is then translated into a sequence of amino acids characteristic of a specific polypeptide.

Restriction Endonuclease. A restriction endonuclease (also restriction enzyme) is an enzyme that has the capacity to recognize a specific base sequence (usually 4, 5, or 6 base pairs in length) in a DNA molecule, and to cleave the DNA molecule at every place where this sequence appears. For example, *Eco*RI recognizes the base sequence GAATTC/CTTAAG.

Restriction Fragment. The DNA molecules produced by digestion with a restriction endonuclease are referred to as restriction fragments. Any given genome can be digested by a particular restriction endonuclease into a discrete set of restriction fragments.

Agarose Gel Electrophoresis. To detect a polymorphism in the length of restriction fragments, an analytical method for fractionating double-stranded DNA molecules on the basis of size is required. The most commonly used technique (though not the only one) for achieving such a fractionation is agarose gel electrophoresis. The principle of this method is that DNA molecules migrate through the gel as though it were a sieve that retards the movement of the largest molecules to the greatest extent and the movement of the smallest molecules to the least extent. Note that the smaller the DNA fragment, the greater the mobility under electrophoresis in the agarose gel.

The DNA fragments fractionated by agarose gel electrophoresis can be visualized directly by a staining procedure if the number of fragments included in the pattern is small. The DNA fragments of genomes can be visualized successfully. However, most genomes, including the human genome, contain far too many DNA sequences to produce a simple pattern of restriction fragments. For example, the human genome is digested into approximately 1,000,000 different DNA fragments by *Eco*RI. In order to visualize a small subset of these fragments, a methodology referred to as the Southern hybridization procedure can be applied.

Southern Transfer Procedure. The purpose of the Southern transfer procedure (also referred to as blotting) is to physically transfer DNA fractionated by agarose gel electrophoresis onto a nitrocellulose filter paper or another appropriate surface or method, while retaining the relative positions of DNA fragments resulting from the fractionation procedure. The methodology used to accomplish the transfer from agarose gel to nitrocellulose involves drawing the DNA from the gel into the nitrocellulose paper by capillary action.

Nucleic Acid Hybridization. Nucleic acid hybridization depends on the principle that two single-stranded nucleic acid molecules that have complementary base sequences will reform the thermodynamically favored double-stranded structure if they are mixed under the proper conditions. The double-stranded structure will be formed between two complementary single-stranded nucleic acids even if one is immobilized on a nitrocellulose filter. In the Southern hybridization procedure, the latter situation occurs. As noted previously, the DNA of the individual to be tested is digested with a restriction endonuclease, fractionated by agarose gel electrophoresis, converted to the single-stranded form, and transferred to nitrocellulose paper, making it available for reannealing to the hybridization probe. Examples of hybridization conditions can be found in Ausubel, F.M. *et al.*, *Current protocols in Molecular Biology*, John Wily & Sons, Inc., New York, NY (1989). A nitrocellulose filter is incubated overnight at 68°C with labeled probe in a solution containing 50% formamide, high salt (either 5x SSC[20X: 3M NaCl/0.3M trisodium citrate] or 5X SSPE [20X: 3.6M NaCl/0.2M NaH_2PO_4/0.02M EDTA, pH 7.7]), 5X Denhardt's solution, 1% SDS, and 100 µg/ml denatured salmon sperm DNA. This is followed by several washes in 0.2X SSC/0.1% SDS at a temperature selected based on the desired stringency: room temperature (low stringency), 42°C (moderate stringency) or 68°C (high stringency). The temperature selected is determined based on the melting temperature (Tm) of the DNA hybrid.

Hybridization Probe. To visualize a particular DNA sequence in the Southern hybridization procedure, a labeled DNA molecule or hybridization

probe is reacted to the fractionated DNA bound to the nitrocellulose filter. The areas on the filter that carry DNA sequences complementary to the labeled DNA probe become labeled themselves as a consequence of the reannealing reaction. The areas of the filter that exhibit such labeling are visualized. The hybridization probe is generally produced by molecular cloning of a specific DNA sequence.

Oligonucleotide or Oligomer. A molecule comprised of two or more deoxyribonucleotides or ribonucleotides, preferably more than three. Its exact size will depend on many factors, which in turn depend on the ultimate function or use of the oligonucleotide. An oligonucleotide can be derived synthetically or by cloning.

Sequence Amplification. A method for generating large amounts of a target sequence. In general, one or more amplification primers are annealed to a nucleic acid sequence. Using appropriate enzymes, sequences found adjacent to, or in between the primers are amplified.

Amplification Primer. An oligonucleotide which is capable of annealing adjacent to a target sequence and serving as an initiation point for DNA synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is initiated.

Vector. A plasmid or phage DNA or other DNA sequence into which DNA can be inserted to be cloned. The vector can replicate autonomously in a host cell, and can be further characterized by one or a small number of endonuclease recognition sites at which such DNA sequences can be cut in a determinable fashion and into which DNA can be inserted. The vector can further contain a marker suitable for use in the identification of cells transformed with the vector. Markers, for example, are tetracycline resistance or ampicillin resistance. The words "cloning vehicle" are sometimes used for "vector."

Expression. Expression is the process by which a structural gene produces a polypeptide. It involves transcription of the gene into mRNA, and the translation of such mRNA into polypeptide(s).

Expression Vector. A vector or vehicle similar to a cloning vector but which is capable of expressing a gene which has been cloned into it, after transformation into a host. The cloned gene is usually placed under the control of (i.e., operably linked to) certain control sequences such as promoter sequences. Expression control sequences will vary depending on whether the vector is designed to express the operably linked gene in a prokaryotic or eukaryotic host and can additionally contain transcriptional elements such as enhancer elements, termination sequences, tissue-specificity elements, and/or translational initiation and termination sites.

Functional Derivative. A "functional derivative" of a sequence, either protein or nucleic acid, is a molecule that possesses a biological activity (either functional or structural) that is substantially similar to a biological activity of the protein or nucleic acid sequence. A functional derivative of a protein can contain post-translational modifications such as covalently linked carbohydrate, depending on the necessity of such modifications for the performance of a specific function. The term "functional derivative" is intended to include the "fragments," "segments," "variants," "analogs," or "chemical derivatives" of a molecule.

As used herein, a molecule is said to be a "chemical derivative" of another molecule when it contains additional chemical moieties not normally a part of the molecule. Such moieties can improve the molecule's solubility, absorption, biological half life, and the like. The moieties can alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, and the like. Moieties capable of mediating such effects are disclosed in *Remington's Pharmaceutical Sciences* (1980). Procedures for coupling such moieties to a molecule are well known in the art.

Variant. A "variant" of a protein or nucleic acid is meant to refer to a molecule substantially similar in structure and biological activity to either the protein or nucleic acid. Thus, provided that two molecules possess a common activity and can substitute for each other, they are considered variants as that term

is used herein even if the composition or secondary, tertiary, or quaternary structure of one of the molecules is not identical to that found in the other, or if the amino acid or nucleotide sequence is not identical.

Allele. An "allele" is an alternative form of a gene occupying a given locus on the chromosome.

Mutation. A "mutation" is any detectable change in the genetic material which can be transmitted to daughter cells and possibly even to succeeding generations giving rise to mutant cells or mutant individuals. If the descendants of a mutant cell give rise only to somatic cells in multicellular organisms, a mutant spot or area of cells arises. Mutations in the germ line of sexually reproducing organisms can be transmitted by the gametes to the next generation resulting in an individual with the new mutant condition in both its somatic and germ cells. A mutation can be any (or a combination of) detectable, unnatural change affecting the chemical or physical constitution, mutability, replication, phenotypic function, or recombination of one or more deoxyribonucleotides; nucleotides can be added, deleted, substituted for, inverted, or transposed to new positions with and without inversion. Mutations can occur spontaneously and can be induced experimentally by application of mutagens. A mutant variation of a nucleic acid molecule results from a mutation. A mutant polypeptide can result from a mutant nucleic acid molecule.

Species. A "species" is a group of actually or potentially interbreeding natural populations. A species variation within a nucleic acid molecule or protein is a change in the nucleic acid or amino acid sequence that occurs among species and can be determined by DNA sequencing of the molecule in question.

Polyacrylamide Gel Electrophoresis (PAGE). The most commonly used technique (though not the only one) for achieving a fractionation of polypeptides on the basis of size is polyacrylamide gel electrophoresis. The principle of this method is that polypeptide molecules migrate through the gel as though it were a sieve that retards the movement of the largest molecules to the greatest extent and the movement of the smallest molecules to the least extent. Note that the

smaller the polypeptide fragment, the greater the mobility under electrophoresis in the polyacrylamide gel. Both before and during electrophoresis, the polypeptides typically are continuously exposed to the detergent sodium dodecyl sulfate (SDS), under which conditions the polypeptides are denatured. Native gels are run in the absence of SDS. The polypeptides fractionated by polyacrylamide gel electrophoresis can be visualized directly by a staining procedure if the number of polypeptide components is small.

Western Transfer Procedure. The purpose of the Western transfer procedure (also referred to as blotting) is to physically transfer polypeptides fractionated by polyacrylamide gel electrophoresis onto a nitrocellulose filter paper or another appropriate surface or method, while retaining the relative positions of polypeptides resulting from the fractionation procedure. The blot is then probed with an antibody that specifically binds to the polypeptide of interest.

Purified. A "purified" protein or nucleic acid is a protein or nucleic acid that has been separated from a cellular component. "Purified" proteins or nucleic acids have been purified to a level of purity not found in nature.

Substantially Pure. A "substantially pure" protein or nucleic acid is a protein or nucleic acid preparation that is lacking in all other cellular components.

Brief Description of the Figures

FIGURE 1. Cosmid contig and transcript map across DYT1 target region. Horizontal lines depict cosmids, vertical lines denote restriction sites X=XhoI, E=EcoRI, and N=NotI (Nv,Xv,Ev = vector ends) and numbers indicate size of fragments in kilobases. The dotted vertical lines indicate the 5' and 3' ends of each cDNA and the position of the recombinant markers D9S2161 and D9S63. Dark horizontal arrows at the bottom represent transcripts, with the direction of the arrow indicating the direction of transcription. The boxes above the cDNAs point out some of the trapped exons. The cosmids used to construct

the map are listed on both sides. The cosmid names followed by LL are from the Lawrence Livermore chromosome 9-specific library and designated names are all preceded by LL09NC01; those followed by LA are from the Los Alamos chromosome 9-specific library. CEN stands for centromere and TEL for telomere and indicate the orientation of this map with respect to chromosome 9q.

FIGURE 2. Sequence variations in cDNAs in DYT1 region. Diagrams are scaled representations of each of the four cDNA transcripts in the critical region. The striped black box indicates 5' untranslated sequence; the striped white box indicates deduced open reading frame; and the wavy line indicates 3' untranslated sequences. The dashed lines at the 5' end of the open reading frame of cDNAs DQ2 and DQ3 indicate that no stop codon 5' to the first predicted methionine (M) has been found. The numbers flanking the open reading frame box indicate the beginning and end of the cDNAs and the nucleotide position of the predicted start and stop codons. Regions generating SSCP shifts are indicated above the transcript diagram: + marks those for which nucleotide changes have not yet been determined; * marks the location of known nucleotide changes corresponding to SSCP shifts; ^ marks the GAG-deletion in cDNA DQ2. Nucleotide changes and resulting amino acid conversions are given below each cDNA. Transcript sizes were estimated by northern blot analysis.

FIGURE 3. Northern blot analysis of DQ2 and DQ1 transcripts. Northern blots of human RNA (Clontech) in order from left to right - fetal: 1) brain, 2) lung, 3) liver and 4) kidney; and adult: 1) heart, 2) brain, 3) placenta, 4) lung, 5) liver, 6) skeletal muscle, 7) kidney and 8) pancreas. Blots hybridized to PCR probes corresponding in Panel A to nucleotides 149 - 1307 of cDNA DQ2; and in Panel B to nucleotides 28 - 728 of cDNA DQ1. Marker sizes are indicated by bars.

FIGURE 4. Comparison of predicted amino acid sequences of torsin gene family members.

A. Alignment of torsins and torps. TorsinA and torsinB are encoded by cDNAs DQ2 and DQ1, respectively. TorpCel is the predicted amino acid sequence from a *C. elegans* genomic sequence. Torp-1 and torp-2 correspond to overlapping expressed sequence tag cDNAs from human and mouse, respectively. The solid triangle represents the site of the GAG (E) deletion in torsinA. Conserved cysteine residues are represented by *. Darkly shaded residues are identical to a consensus sequence; lightly shaded residues are similar. Conserved possible phosphorylation sites for protein kinase C (PKC) and casein kinase 2 (CK2) are boxed.

B. Schematic representation of torsinA domains. The N-terminal region (left) contains about 40 hydrophobic amino acids, preceded by two basic residues (K, R) and bisected by a polar and an acidic residue (Q, E). The ATP-binding domain is indicated along with its conserved A and B motifs. Two additional motifs conserved with the HSP100 family (SN and IV) are shaded.

C. Comparison of torsins and torps with two representative members of the HSP100 family. SKD3, from mouse is an HSP100 family member of class 2M; HSP101 from soybean is a heat shock protein of class 1B (Schirmer, E. C., *et al.*, *TIBS 21*:289-296 (1996)). Shaded residues are identical to a consensus sequence. The conserved motifs (A, B, SN and IV) occur in all seven proteins.

FIGURE 5. Resolution of GAG-deletion associated with early onset dystonia.

A. Sequence. Autoradiograph of sequencing gel showing mutant and normal sequence from amplified genomic DNA using primer 6419. The sequence, which is read 5' to 3' from the bottom up, demonstrates the GAG-deletion found on one allele in an affected patient. This mutation occurs at nucleotides 946-948 in the coding region and results in the deletion of a Glu residue from the protein.

B. SSCP analysis of fragments. Autoradiograph showing PCR-SSCP analysis of genomic DNA from affected and control individuals. Primers (6418 and 6419) were used to produce 250 bp fragment which were resolved in non-denaturing acrylamide gels. Lane 1, 3, and 5 are affected individuals with typical early onset dystonia; lane 2 and 4 are unaffected individuals. Solid triangle indicates shifted band associated with GAG-deletion; open triangle indicates 2 bands associated with normal allele.

C. Digestion of PCR fragment with BseRI. PCR products were generated from genomic DNA with primers 6419 and H48 and the 200 bp product was digested with BseRI. Bands of 120 bp and 70 bp were generated from control DNA (lane 2), whereas a novel band of 130 bp was generated from affected individuals with the GAG-deletion (lane 1), because of the loss of a BseRI site. Lane 3 and 4 are markers: 3) PCR products of specific sizes (50, 100 and 130 bp) and 4) 100 bp ladder (Pharmacia).

D. Sequence surrounding GAG-deletion. Normal genomic/cDNA sequence of torsinA showing position of primers (arrows); the GAGGAG sequence (bracketed) in which the GAG-deletion occurs; and BseRI sites (*, site deletion lost as a result of the deletion).

FIGURE 6. Hypothetical model of neuronal involvement in early onset dystonia as a function of age. Dystonic symptoms are believed to arise through neuronal dysfunction in the basal ganglia. Neurons in this region of the brain have a anatomical patterning corresponding to the movements they subserve in the body, which can be represented roughly as an inverted homunculus. In early onset dystonia the physical site of onset of symptoms and tendency to generalize as a function of age can be represented as a zone of susceptibility (cross-hatched area) moving ventrally in the basal ganglia with age. Clinical analysis of carriers of the AJ founder mutation (Bressman, S. B., *et al., Annal Neurol 36*:771-777 (1994b)) reveals that the earlier the onset of symptoms, the more likely they are to commence in lower limbs and the greater the tendency (thick black arrows) to

generalize and involve upper parts of the body. With increasing age of onset, symptoms tend to involve progressively higher body parts, and still tend to progress upward. By >28 years of age, carriers of the early onset gene have passed the age of susceptibility and have only a small remaining chance of manifesting any symptoms. Still, as gene carriers, these "escapees" are at equal risk as affected gene carriers for having affected children.

Detailed Description of the Preferred Embodiments

For purposes of clarity of disclosure, and not by way of limitation, the detailed description of the invention is divided into the following subsections:

I. Isolated Nucleic Acid Molecules Coding for Torsin Polypeptides.
II. Purified Torsin Polypeptides.
III. A Nucleic Acid Probe for the Specific Detection of Torsin Nucleic Acid.
IV. A Method of Detecting the Presence of Torsin Nucleic Acid in a Sample.
V. A Kit for Detecting the Presence of Torsin Nucleic Acid in a Sample.
VI. DNA Constructs Comprising a Torsin Nucleic Acid Molecule and Cells Containing These Constructs.
VII. An Antibody Having Binding Affinity to a Torsin Polypeptide and a Hybridoma Containing the Antibody.
VIII. A Method of Detecting a Torsin Polypeptide or Antibody in a Sample.
IX. A Diagnostic Kit Comprising a Torsin Protein or Antibody.
X. Diagnostic Screening and Treatment
XI. Transgenic Torsin "Knock-out" Mice
XII. HSV-1 Amplicon Constructs

I. Isolated Nucleic Acid Molecules Coding for Torsin Polypeptides

In one embodiment, the present invention relates to isolated nucleic acid molecules comprising a polynucleotide sequence at least 90% identical (more preferably, 95%, 96%, 97%, 98%, 99% or 100% identical) to a sequence selected from the group consisting of:

(a) a nucleotide sequence encoding the torsin polypeptide comprising the complete amino acid sequence in SEQ ID NO: 2 or 4;

(b) a nucleotide sequence encoding the torsin polypeptide comprising the complete amino acid sequence encoded by the polynucleotide clone contained in ATCC Deposit No. 98454 or 98455; and

(c) a nucleotide sequence complementary to any of the nucleotide sequences in (a) or (b).

The torsin nucleic acids were deposited at the American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD 20852, USA on June 12, 1997 as ATCC Nos. 98454 and 98455.

In one preferred embodiment, the isolated nucleic acid molecule comprises a torsin nucleotide sequence with greater than 90% identity or similarity to the nucleotide sequence present in SEQ ID NO: 1 or 3 (preferably greater than 95%, 96%, 97%, 98%, 99% or 100%). In another preferred embodiment, the isolated nucleic acid molecule comprises the torsin nucleotide sequence present in SEQ ID NO:1 or 3. In another embodiment, the isolated nucleic acid molecule encodes the torsin amino acid sequence present in SEQ ID NO: 2 or 4.

Also included within the scope of this invention are the functional equivalents of the herein-described isolated nucleic acid molecules and derivatives thereof. For example, the nucleic acid sequences depicted in SEQ ID NO:1 or 3 can be altered by substitutions, additions or deletions that provide for functionally equivalent molecules. Due to the degeneracy of nucleotide coding sequences, other DNA sequences which encode substantially the same amino acid sequence as depicted in SEQ ID NO: 2 or 4 can be used in the practice of the

present invention. These include but are not limited to nucleotide sequences comprising all or portions of torsin nucleic acid depicted in SEQ ID NO:1 or 3 which are altered by the substitution of different codons that encode a functionally equivalent amino acid residue within the sequence.

In addition, the nucleic acid sequence can comprise a nucleotide sequence which results from the addition, deletion or substitution of at least one nucleotide to the 5'-end and/or the 3'-end of the nucleic acid formula shown in SEQ ID NO:1 or 3 or a derivative thereof. Any nucleotide or polynucleotide can be used in this regard, provided that its addition, deletion or substitution does not substantially alter the amino acid sequence of SEQ ID NO:2 or 4 which is encoded by the nucleotide sequence. Moreover, the nucleic acid molecule of the present invention can, as necessary, have restriction endonuclease recognition sites added to its 5'-end and/or 3'-end. All variations of the nucleotide sequence of the torsin gene and fragments thereof permitted by the genetic code are, therefore, included in this invention.

Further, it is possible to delete codons or to substitute one or more codons by codons other than degenerate codons to produce a structurally modified polypeptide, but one which has substantially the same utility or activity of the polypeptide produced by the unmodified nucleic acid molecule. As recognized in the art, the two polypeptides are functionally equivalent, as are the two nucleic acid molecules which give rise to their production, even though the differences between the nucleic acid molecules are not related to degeneracy of the genetic code.

A. Isolation of Nucleic Acid

In one aspect of the present invention, isolated nucleic acid molecules coding for polypeptides having amino acid sequences corresponding to torsin are provided. In particular, the nucleic acid molecule can be isolated from a biological sample containing torsin RNA or DNA.

The nucleic acid molecule can be isolated from a biological sample containing torsin RNA using the techniques of cDNA cloning and subtractive hybridization. The nucleic acid molecule can also be isolated from a cDNA library using a homologous probe.

The nucleic acid molecule can be isolated from a biological sample containing genomic DNA or from a genomic library. Suitable biological samples include, but are not limited to, whole organisms, organs, tissues, blood and cells. The method of obtaining the biological sample will vary depending upon the nature of the sample.

One skilled in the art will realize that genomes can be subject to slight allelic variations between individuals. Therefore, the isolated nucleic acid molecule is also intended to include allelic variations, so long as the sequence is a functional derivative of the torsin coding sequence. When a torsin allele does not encode the identical sequence to that found in SEQ ID NO:1 or 3, it can be isolated and identified as torsin using the same techniques used herein, and especially PCR techniques to amplify the appropriate gene with primers based on the sequences disclosed herein.

One skilled in the art will realize that organisms other than humans will also contain torsin genes (for example, eukaryotes; more specifically, mammals, rodents, worms (preferably, *C. elegans*), insects (preferably, fruit flies, *Drosophila*) birds, fish, yeast, and plants; more specifically, gorillas, rhesus monkeys, and chimpanzees). The invention is intended to include, but not be limited to, torsin nucleic acid molecules isolated from the above-described organisms.

B. Synthesis of Nucleic Acid

Isolated nucleic acid molecules of the present invention are also meant to include those chemically synthesized. For example, a nucleic acid molecule with the nucleotide sequence which codes for the expression product of an torsin gene

can be designed and, if necessary, divided into appropriate smaller fragments. Then an oligomer which corresponds to the nucleic acid molecule, or to each of the divided fragments, can be synthesized. Such synthetic oligonucleotides can be prepared, for example, by the triester method of Matteucci *et al.*, *J. Am. Chem. Soc. 103*:3185-3191 (1981) or by using an automated DNA synthesizer.

An oligonucleotide can be derived synthetically or by cloning. If necessary, the 5'-ends of the oligomers can be phosphorylated using T4 polynucleotide kinase. Kinasing of single strands prior to annealing or for labeling can be achieved using an excess of the enzyme. If kinasing is for the labeling of probe, the ATP can contain high specific activity radioisotopes. Then, the DNA oligomer can be subjected to annealing and ligation with T4 ligase or the like.

II. Purified Torsin Polypeptides

In another embodiment, the present invention relates to a purified polypeptide (preferably, substantially pure) having an amino acid sequence corresponding to torsin, or a functional derivative thereof. In a preferred embodiment, the polypeptide has the amino acid sequence set forth in SEQ ID NO: 2 or 4 or mutant or species variation thereof, or at least 80% identity or at least 90% similarity thereof (preferably, at least 90%, 95%, 96%, 97%, 98%, or 99% identity or at least 95%, 96%, 97%, 98%, or 99% similarity thereof), or at least 6 contiguous amino acids thereof (preferably, at least 10, 15, 20, 25, or 50 contiguous amino acids thereof).

In a preferred embodiment, the invention relates to torsin epitopes. The epitope of these polypeptides is an immunogenic or antigenic epitope. An immunogenic epitope is that part of the protein which elicits an antibody response when the whole protein is the immunogen. An antigenic epitope is a fragment of the protein which can elicit an antibody response. Methods of selecting antigenic epitope fragments are well known in the art. *See,* Sutcliffe *et al.*,

*Science 219:*660-666 (1983). Antigenic epitope-bearing peptides and polypeptides of the invention are useful to raise an immune response that specifically recognizes the polypeptides. Antigenic epitope-bearing peptides and polypeptides of the invention comprise at least 7 amino acids (preferably, 9, 10, 12, 15 or 20 amino acids) of the proteins of the invention. Examples of antigenic polypeptides or peptides include those listed in Table 1, below.

TABLE 1: ANTIGENIC EPITOPES

	Size(Number of Amino Acids)	AA Position
Torsin A	5	50-55
	5	92-96
	6	225-230
	7	287-293
	6	315-320
Torsin B	9	49-57
	3	133-135
	4	189-192
	7	275-281

Amino acid sequence variants of torsin can be prepared by mutations in the DNA. Such variants include, for example, deletions from, or insertions or substitutions of, residues within the amino acid sequence shown in SEQ ID NO:2 or 4. Any combination of deletion, insertion, and substitution can also be made to arrive at the final construct, provided that the final construct possesses the desired activity.

While the site for introducing an amino acid sequence variation is predetermined, the mutation *per se* need not be predetermined. For example, to optimize the performance of a mutation at a given site, random mutagenesis can be conducted at the target codon or region and the expressed torsin variants

screened for the optimal combination of desired activity. Techniques for making substitution mutations at predetermined sites in DNA having a known sequence are well known, for example, site-specific mutagenesis.

Preparation of a torsin variant in accordance herewith is preferably achieved by site-specific mutagenesis of DNA that encodes an earlier prepared variant or a nonvariant version of the protein. Site-specific mutagenesis allows the production of torsin variants through the use of specific oligonucleotide sequences that encode the DNA sequence of the desired mutation. In general, the technique of site-specific mutagenesis is well known in the art, as exemplified by publications such as Adelman *et al.*, *DNA 2*:183 (1983) and Ausubel *et al.* "Current Protocols in Molecular Biology", J. Wiley & Sons, NY, NY, 1996.

Amino acid sequence deletions generally range from about 1 to 30 residues, more preferably 1 to 10 residues, and typically are contiguous.

Amino acid sequence insertions include amino and/or carboxyl-terminal fusions of from one residue to polypeptides of essentially unrestricted length, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions (i.e., insertions within the complete torsin sequence) can range generally from about 1 to 10 residues, more preferably 1 to 5.

The third group of variants are those in which at least one amino acid residue in the torsin molecule, and preferably, only one, has been removed and a different residue inserted in its place. Such substitutions preferably are made in accordance with the following Table 2 when it is desired to modulate finely the characteristics of torsin.

TABLE 2

Original Residue	Exemplary Substitutions
Ala	gly; ser
Arg	lys
Asn	gln; his
Asp	glu
Cys	ser
Gln	asn
Glu	asp
Gly	ala; pro
His	asn; gln
Ile	leu; val
Leu	ile; val
Lys	arg; gln; glu
Met	leu; tyr; ile
Phe	met; leu; tyr
Ser	thr
Thr	ser
Trp	tyr
Tyr	trp; phe
Val	ile; leu

Substantial changes in functional or immunological identity are made by selecting substitutions that are less conservative than those in Table 2, i.e., selecting residues that differ more significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain. The substitutions that in general are expected are those in which (a) glycine and/or proline is substituted by another amino acid or is deleted or inserted; (b) a hydrophilic residue, e.g., seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g., leucyl, isoleucyl, phenylalanyl, valyl, or alanyl; (c) a cysteine residue is substituted for (or by) any other residue; (d) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histidyl, is substituted for (or by) a residue having an electronegative charge, e.g., glutamyl or aspartyl; or (e) a residue

having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having such a side chain, e.g., glycine.

Some deletions and insertions, and substitutions are not expected to produce radical changes in the characteristics of torsin. However, when it is difficult to predict the exact effect of the substitution, deletion, or insertion in advance of doing so, one skilled in the art will appreciate that the effect will be evaluated by routine screening assays. For example, a variant typically is made by site-specific mutagenesis of the native torsin encoding-nucleic acid, expression of the variant nucleic acid in recombinant cell culture, and, optionally, purification from the cell culture, for example, by immunoaffinity adsorption on a column (to absorb the variant by binding it to at least one remaining immune epitope). The activity of the cell lysate or purified torsin molecule variant is then screened in a suitable screening assay for the desired characteristic. For example, a change in the immunological character of the torsin molecule, such as affinity for a given antibody, is measured by a competitive type immunoassay. Changes in immunomodulation activity are measured by the appropriate assay. Modifications of such protein properties as redox or thermal stability, hydrophobicity, susceptibility to proteolytic degradation or the tendency to aggregate with carriers or into multimers are assayed by methods well known to the ordinarily skilled artisan.

A variety of methodologies known in the art can be utilized to obtain the peptide of the present invention. In one embodiment, the peptide is purified from tissues or cells which naturally produce the peptide. Alternatively, the above-described isolated nucleic acid fragments can be used to express the torsin protein in any organism. The samples of the present invention include cells, protein extracts or membrane extracts of cells, or biological fluids. The sample will vary based on the assay format, the detection method and the nature of the tissues, cells or extracts used as the sample.

Any organism can be used as a source for the peptide of the invention, as long as the source organism naturally contains such a peptide. As used herein,

"source organism" refers to the original organism from which the amino acid sequence of the subunit is derived, regardless of the organism the subunit is expressed in and ultimately isolated from.

One skilled in the art can readily follow known methods for isolating proteins in order to obtain the peptide free of natural contaminants. These include, but are not limited to: immunochromotography, size-exclusion chromatography, HPLC, ion-exchange chromatography, and immuno-affinity chromatography.

In a preferred embodiment, the purification procedures comprise ion-exchange chromatography and size exclusion chromatography. Any one of a large number of ion-exchange resins known in the art can be employed, including for example, monoQ, sepharose Q, macro-prepQ, AG1-X2, or HQ. Examples of suitable size exclusion resins include, but are not limited to, Superdex 200, Superose 12, and Sephycryl 200. Elution can be achieved with aqueous solutions of potassium chloride or sodium chloride at concentrations ranging from 0.01M to 2.0M.

III. A Nucleic Acid Probe for the Specific Detection of Torsin Nucleic Acid

In another embodiment, the present invention relates to a nucleic acid probe for the specific detection of the presence of torsin nucleic acid in a sample comprising the above-described nucleic acid molecules or at least a fragment thereof which binds under stringent conditions to torsin nucleic acid.

In one prefered embodiment, the present invention relates to an isolated nucleic acid probe consisting of 10 to 1000 nucleotides (prefererably, 10 to 500, 10 to 100, 10 to 50, 10 to 35, 20 to 1000, 20 to 500, 20 to 100, 20 to 50, or 20 to 35) which hybridizes preferentially to RNA or DNA of torsin but not to RNA or DNA of which is not related to torsin, wherein said nucleic acid probe is or is complementary to a nucleotide sequence consisting of at least 10 consecutive nucleotides (preferably, 15, 18, 20, 25, or 30) from the nucleic acid molecule

comprising a polynucleotide sequence at least 90% identical to a sequence selected from the group consisting of:

(a) a nucleotide sequence encoding the torsin polypeptide comprising the complete amino acid sequence in SEQ ID NO: 2 or 4;

(b) a nucleotide sequence encoding the torsin polypeptide comprising the complete amino acid sequence encoded by the polynucleotide clone contained in ATCC Deposit No. 98454 or 98455;

(c) a nucleotide sequence complementary to any of the nucleotide sequences in (a) or (b) and

(d) a nucleotide sequence as previously described above.

Examples of specific nucleic acid probes which can be used in the present invention are set forth in Table 3, below.

TABLE 3: NUCLEIC ACID PROBES

Torsin A	Size (no. of bases)	Nucleotides
	20	43-62 (SEQ ID NO:5)
	20	63-82 (SEQ ID NO:5)
	40	43-82 (SEQ ID NO:5)
	100	43-142 (SEQ ID NO:5)
	100	143-242 (SEQ ID NO:5)
	1158	149-1307 (SEQ ID NO:5)
Torsin B	20	994-1013 (SEQ ID NO:3)
	20	1014-1033 (SEQ ID NO:3)
	40	994-1033 (SEQ ID NO:3)
	100	994-1093 (SEQ ID NO:3)
	100	1094-1193 (SEQ ID NO:3)
	700	28-728 (SEQ ID NO:6)

The nucleic acid probe can be used to probe an appropriate chromosomal or cDNA library by usual hybridization methods to obtain another nucleic acid

molecule of the present invention. A chromosomal DNA or cDNA library can be prepared from appropriate cells according to recognized methods in the art (cf. *Molecular Cloning: A Laboratory Manual, second edition*, edited by Sambrook, Fritsch, & Maniatis, Cold Spring Harbor Laboratory, 1989).

In the alternative, chemical synthesis is carried out in order to obtain nucleic acid probes having nucleotide sequences which correspond to N-terminal and C-terminal portions of the torsin amino acid sequence (See, Table 3). Thus, the synthesized nucleic acid probes can be used as primers in a polymerase chain reaction (PCR) carried out in accordance with recognized PCR techniques, essentially according to *PCR Protocols, A Guide to Methods and Applications*, edited by Michael *et al.*, Academic Press, 1990, utilizing the appropriate chromosomal, cDNA or cell line library to obtain the fragment of the present invention.

One skilled in the art can readily design such probes based on the sequence disclosed herein using methods of computer alignment and sequence analysis known in the art (cf. *Molecular Cloning: A Laboratory Manual, second edition*, edited by Sambrook, Fritsch, & Maniatis, Cold Spring Harbor Laboratory, 1989).

The hybridization probes of the present invention can be labeled by standard labeling techniques such as with a radiolabel, enzyme label, fluorescent label, biotin-avidin label, chemiluminescence, and the like. After hybridization, the probes can be visualized using known methods.

The nucleic acid probes of the present invention include RNA, as well as DNA probes, such probes being generated using techniques known in the art.

In one embodiment of the above described method, a nucleic acid probe is immobilized on a solid support. Examples of such solid supports include, but are not limited to, plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, and acrylic resins, such as polyacrylamide and latex beads. Techniques for coupling nucleic acid probes to such solid supports are well known in the art.

The test samples suitable for nucleic acid probing methods of the present invention include, for example, cells or nucleic acid extracts of cells, or biological fluids. The sample used in the above-described methods will vary based on the assay format, the detection method and the nature of the tissues, cells or extracts to be assayed. Methods for preparing nucleic acid extracts of cells are well known in the art and can be readily adapted in order to obtain a sample which is compatible with the method utilized.

IV. A Method of Detecting The Presence of Torsin Nucleic Acid in a Sample

In another embodiment, the present invention relates to a method of detecting the presence of torsin nucleic acid in a sample comprising a) contacting the sample with the above-described nucleic acid probe, under specific hybridization conditions such that hybridization occurs, and b) detecting the presence of the probe bound to the nucleic acid molecule. One skilled in the art would select the nucleic acid probe according to techniques known in the art as described above. Samples to be tested include but should not be limited to RNA or DNA samples from human tissue.

V. A Kit for Detecting the Presence of Torsin Nucleic Acid in a Sample

In another embodiment, the present invention relates to a kit for detecting the presence of torsin nucleic acid in a sample comprising at least one container means having disposed therein the above-described nucleic acid probe. In a preferred embodiment, the kit further comprises other containers comprising one or more of the following: wash reagents and reagents capable of detecting the presence of bound nucleic acid probe. Examples of detection reagents include, but are not limited to radiolabelled probes, enzymatic labeled probes (horse radish peroxidase, alkaline phosphatase), and affinity labeled probes (biotin, avidin, or steptavidin).

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allow the efficient transfer of reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the probe or primers used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, and the like), and containers which contain the reagents used to detect the hybridized probe, bound antibody, amplified product, or the like.

One skilled in the art will readily recognize that the nucleic acid probes described in the present invention can readily be incorporated into one of the established kit formats which are well known in the art.

VI. DNA Constructs Comprising a Torsin Nucleic Acid Molecule and Cells Containing These Constructs

In another embodiment, the present invention relates to a recombinant DNA molecule comprising, 5′ to 3′, a promoter effective to initiate transcription in a host cell and the above-described nucleic acid molecules. In another embodiment, the present invention relates to a recombinant DNA molecule comprising a vector and an above-described nucleic acid molecule.

In another embodiment, the present invention relates to a nucleic acid molecule comprising a transcriptional control region functional in a cell, a sequence complimentary to an RNA sequence encoding an amino acid sequence corresponding to the above-described polypeptide, and a transcriptional termination region functional in the cell.

Preferably, the above-described molecules are isolated and/or purified DNA molecules.

In another embodiment, the present invention relates to a cell or non-human organism that contains an above-described nucleic acid molecule.

In another embodiment, the peptide is purified from cells which have been altered to express the peptide.

As used herein, a cell is said to be "altered to express a desired peptide" when the cell, through genetic manipulation, is made to produce a protein which it normally does not produce or which the cell normally produces at low levels. One skilled in the art can readily adapt procedures for introducing and expressing either genomic, cDNA, or synthetic sequences into either eukaryotic or prokaryotic cells.

A nucleic acid molecule, such as DNA, is said to be "capable of expressing" a polypeptide if it contains nucleotide sequences which contain transcriptional and translational regulatory information and such sequences are "operably linked" to nucleotide sequences which encode the polypeptide. An operable linkage is a linkage in which the regulatory DNA sequences and the DNA sequence sought to be expressed are connected in such a way as to permit gene sequence expression. The precise nature of the regulatory regions needed for gene sequence expression can vary from organism to organism, but shall in general include a promoter region which, in prokaryotes, contains both the promoter (which directs the initiation of RNA transcription) as well as the DNA sequences which, when transcribed into RNA, will signal synthesis initiation. Such regions will normally include those 5′-non-coding sequences involved with initiation of transcription and translation, such as the TATA box, capping sequence, CAAT sequence, and the like.

If desired, the non-coding region 3′ to the torsin coding sequence can be obtained by the above-described methods. This region can be retained for its transcriptional termination regulatory sequences, such as termination and polyadenylation. Thus, by retaining the 3′-region naturally contiguous to the DNA sequence encoding an torsin gene, the transcriptional termination signals can be provided. Where the transcriptional termination signals are not

satisfactorily functional in the expression host cell, then a 3′ region functional in the host cell can be substituted.

Two DNA sequences (such as a promoter region sequence and an torsin coding sequence) are said to be operably linked if the nature of the linkage between the two DNA sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter region sequence to direct the transcription of a torsin coding sequence, or (3) interfere with the ability of the torsin coding sequence to be transcribed by the promoter region sequence. Thus, a promoter region would be operably linked to a DNA sequence if the promoter were capable of effecting transcription of that DNA sequence.

The present invention encompasses the expression of the torsin coding sequence (or a functional derivative thereof) in either prokaryotic or eukaryotic cells. Prokaryotic hosts are, generally, the most efficient and convenient for the production of recombinant proteins and, therefore, are preferred for the expression of the torsin coding sequence.

Prokaryotes most frequently are represented by various strains of *E. coli*. However, other microbial strains can also be used, including other bacterial strains. In prokaryotic systems, plasmid vectors that contain replication sites and control sequences derived from a species compatible with the host can be used. Examples of suitable plasmid vectors include pBR322, pUC18, pUC19, pUC118, pUC119 and the like; suitable phage or bacteriophage vectors include λgt10, λgt11 and the like; and suitable virus vectors include pMAM-neo, pKRC and the like. Preferably, the selected vector of the present invention has the capacity to replicate in the selected host cell.

Recognized prokaryotic hosts include bacteria such as *E. coli, Bacillus, Streptomyces, Pseudomonas, Salmonella, Serratia,* and the like. However, under such conditions, the peptide will not be glycosylated. The prokaryotic host must be compatible with the replicon and control sequences in the expression plasmid.

To express torsin in a prokaryotic cell, it is necessary to operably link the torsin coding sequence to a functional prokaryotic promoter. Such promoters can be either constitutive or, more preferably, regulatable (i.e., inducible or derepressible). Examples of constitutive promoters include the *int* promoter of bacteriophage λ, the *bla* promoter of the β-lactamase gene sequence of pBR322, and the CAT promoter of the chloramphenicol acetyl transferase gene sequence of pBR325, and the like. Examples of inducible prokaryotic promoters include the major right and left promoters of bacteriophage λ (P_L and P_R), the *trp, recA, lacZ, lacI,* and *gal* promoters of *E. coli*, the α-amylase (Ulmanen *et al., J. Bacteriol. 162*:176-182 (1985)) and the ς-28-specific promoters of *B. subtilis* (Gilman *et al., Gene sequence 32*:11-20 (1984)), the promoters of the bacteriophages of *Bacillus* (Gryczan, In: *The Molecular Biology of the Bacilli*, Academic Press, Inc., NY (1982)), and *Streptomyces* promoters (Ward *et al., Mol. Gen. Genet. 203*:468-478 (1986)). Prokaryotic promoters are reviewed by Glick (*J. Ind. Microbiol. 1*:277-282 (1987)); Cenatiempo (*Biochimie 68*:505-516 (1986)); and Gottesman (*Ann. Rev. Genet. 18*:415-442 (1984)).

Proper expression in a prokaryotic cell also requires the presence of a ribosome binding site upstream of the gene sequence-encoding sequence. Such ribosome binding sites are disclosed, for example, by Gold *et al.* (*Ann. Rev. Microbiol. 35*:365-404 (1981)).

The selection of control sequences, expression vectors, transformation methods, and the like, are dependent on the type of host cell used to express the gene. As used herein, "cell", "cell line", and "cell culture" can be used interchangeably and all such designations include progeny. Thus, the words "transformants" or "transformed cells" include the primary subject cell and cultures derived therefrom, without regard to the number of transfers. It is also understood that all progeny can not be precisely identical in DNA content, due to deliberate or inadvertent mutations. However, as defined, mutant progeny have the same functionality as that of the originally transformed cell.

Host cells which can be used in the expression systems of the present invention are not strictly limited, provided that they are suitable for use in the expression of the torsin peptide of interest. Suitable hosts include eukaryotic cells.

Preferred eukaryotic hosts include, for example, yeast, fungi, insect cells, mammalian cells either *in vivo*, or in tissue culture. Preferred mammalian cells include HeLa cells, cells of fibroblast origin such as VERO or CHO-K1, or cells of lymphoid origin and their derivatives.

In addition, plant cells are also available as hosts, and control sequences compatible with plant cells are available, such as the cauliflower mosaic virus 35S and 19S, and nopaline synthase promoter and polyadenylation signal sequences.

Another preferred host is an insect cell, for example *Drosophila* larvae. Using insect cells as hosts, the *Drosophila* alcohol dehydrogenase promoter can be used, Rubin, *Science 240*:1453-1459 (1988). Alternatively, baculovirus vectors can be engineered to express large amounts of torsin in insect cells (Jasny, *Science 238*:1653 (1987); Miller *et al.*, In: *Genetic Engineering* (1986), Setlow, J.K., *et al.*, eds., *Plenum*, Vol. 8, pp. 277-297).

Different host cells have characteristic and specific mechanisms for the translational and post-translational processing and modification (*e.g.*, glycosylation, cleavage) of proteins. Appropriate cell lines or host systems can be chosen to ensure the desired modification and processing of the foreign protein expressed.

Any of a series of yeast gene sequence expression systems can be utilized which incorporate promoter and termination elements from the actively expressed gene sequences coding for glycolytic enzymes. These enzymes are produced in large quantities when yeast are grown in mediums rich in glucose. Known glycolytic gene sequences can also provide very efficient transcriptional control signals.

Yeast provides substantial advantages in that it can also carry out post-translational peptide modifications. A number of recombinant DNA strategies exist which utilize strong promoter sequences and high copy number of plasmids which can be utilized for production of the desired proteins in yeast. Yeast recognizes leader sequences on cloned mammalian gene sequence products and secretes peptides bearing leader sequences (i.e., pre-peptides). For a mammalian host, several possible vector systems are available for the expression of torsin.

A wide variety of transcriptional and translational regulatory sequences can be employed, depending upon the nature of the host. The transcriptional and translational regulatory signals can be derived from viral sources, such as adenovirus, bovine papilloma virus, simian virus, or the like, where the regulatory signals are associated with a particular gene sequence which has a high level of expression. Alternatively, promoters from mammalian expression products, such as actin, collagen, myosin, and the like, can be employed. Transcriptional initiation regulatory signals can be selected which allow for repression or activation, so that expression of the gene sequences can be modulated. Of interest are regulatory signals which are temperature-sensitive so that by varying the temperature, expression can be repressed or initiated, or are subject to chemical (such as metabolite) regulation.

As discussed above, expression of torsin in eukaryotic hosts requires the use of eukaryotic regulatory regions. Such regions will, in general, include a promoter region sufficient to direct the initiation of RNA synthesis. Preferred eukaryotic promoters include, for example, the promoter of the mouse metallothionein I gene sequence (Hamer *et al.*, *J. Mol. Appl. Gen. 1*:273-288 (1982)); the TK promoter of Herpes virus (McKnight, *Cell 31*:355-365 (1982)); the SV40 early promoter (Benoist *et al.*, *Nature (London) 290*:304-310 (1981)); the yeast *gal4* gene sequence promoter (Johnston *et al.*, *Proc. Natl. Acad. Sci. (USA) 79*:6971-6975 (1982); Silver *et al.*, *Proc. Natl. Acad. Sci. (USA) 81*:5951-5955 (1984)) and the CMV immediate-early gene promoter (Thomsen *et al.*, *Proc. Natl. Acad. Sci (USA) 81*:659-663 (1984).

As is widely known, translation of eukaryotic mRNA is initiated at the codon which encodes the first methionine. For this reason, it is preferable to ensure that the linkage between a eukaryotic promoter and a torsin coding sequence does not contain any intervening codons which are capable of encoding a methionine (i.e., AUG). The presence of such codons results either in a formation of a fusion protein (if the AUG codon is in the same reading frame as the torsin coding sequence) or a frame-shift mutation (if the AUG codon is not in the same reading frame as the torsin coding sequence).

A torsin nucleic acid molecule and an operably linked promoter can be introduced into a recipient prokaryotic or eukaryotic cell either as a non-replicating DNA (or RNA) molecule, which can either be a linear molecule or, more preferably, a closed covalent circular molecule. Since such molecules are incapable of autonomous replication, the expression of the gene can occur through the transient expression of the introduced sequence. Alternatively, permanent expression can occur through the integration of the introduced DNA sequence into the host chromosome.

In one embodiment, a vector is employed which is capable of integrating the desired gene sequences into the host cell chromosome. Cells which have stably integrated the introduced DNA into their chromosomes can be selected by also introducing one or more markers which allow for selection of host cells which contain the expression vector. The marker can provide for prototrophy to an auxotrophic host, biocide resistance, e.g., antibiotics, or heavy metals, such as copper, or the like. The selectable marker gene sequence can either be directly linked to the DNA gene sequences to be expressed, or introduced into the same cell by co-transfection. Additional elements can also be needed for optimal synthesis of single chain binding protein mRNA. These elements can include splice signals, as well as transcription promoters, enhancer signal sequences, and termination signals. cDNA expression vectors incorporating such elements include those described by Okayama, *Molec. Cell. Biol. 3*:280 (1983).

In a preferred embodiment, the introduced nucleic acid molecule will be incorporated into a plasmid or viral vector capable of autonomous replication in the recipient host. Any of a wide variety of vectors can be employed for this purpose. Factors of importance in selecting a particular plasmid or viral vector include: the ease with which recipient cells that contain the vector can be recognized and selected from those recipient cells which do not contain the vector; the number of copies of the vector which are desired in a particular host; and whether it is desirable to be able to "shuttle" the vector between host cells of different species. Preferred prokaryotic vectors include plasmids such as those capable of replication in *E. coli* (such as, for example, pBR322, ColE1, pSC101, pACYC 184, πVX. Such plasmids are, for example, disclosed by Sambrook (cf. *Molecular Cloning: A Laboratory Manual*, second edition, edited by Sambrook, Fritsch, & Maniatis, Cold Spring Harbor Laboratory, 1989). *Bacillus* plasmids include pC194, pC221, pT127, and the like. Such plasmids are disclosed by Gryczan (In: *The Molecular Biology of the Bacilli*, Academic Press, NY (1982), pp. 307-329). Suitable *Streptomyces* plasmids include pIJ101 (Kendall *et al., J. Bacteriol. 169*:4177-4183 (1987)), and streptomyces bacteriophages such as ϕC31 (Chater *et al.*, In: *Sixth International Symposium on Actinomycetales Biology*, Akademiai Kaido, Budapest, Hungary (1986), pp. 45-54). Pseudomonas plasmids are reviewed by John *et al. (Rev. Infect. Dis. 8*:693-704 (1986)), and Izaki (*Jpn. J. Bacteriol. 33*:729-742 (1978)).

Preferred eukaryotic plasmids include, for example, BPV, vaccinia, SV40, 2-micron circle, and the like, or their derivatives. Such plasmids are well known in the art (Botstein *et al., Miami Wntr. Symp. 19*:265-274 (1982); Broach, In: *The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, p. 445-470 (1981); Broach, *Cell 28*:203-204 (1982); Bollon *et al., J. Clin. Hematol. Oncol. 10*:39-48 (1980); Maniatis, In: *Cell Biology: A Comprehensive Treatise*, Vol. 3, Gene Sequence Expression, Academic Press, NY, pp. 563-608 (1980)).

Once the vector or nucleic acid molecule containing the construct(s) has been prepared for expression, the DNA construct(s) can be introduced into an appropriate host cell by any of a variety of suitable means, i.e., transformation, transfection, conjugation, protoplast fusion, electroporation, particle gun technology, calcium phosphate-precipitation, direct microinjection, and the like. After the introduction of the vector, recipient cells are grown in a selective medium, which selects for the growth of vector-containing cells. Expression of the cloned gene molecule(s) results in the production of torsin. This can take place in the transformed cells as such, or following the induction of these cells to differentiate (for example, by administration of bromodeoxyuracil to neuroblastoma cells or the like).

VII. An Antibody Having Binding Affinity to a Torsin Polypeptide and a Hybridoma Containing the Antibody

In another embodiment, the present invention relates to an antibody having binding affinity specifically to a torsin polypeptide as described above or specifically to a torsin polypeptide binding fragment thereof. An antibody binds specifically to a torsin polpeptide or binding fragment thereof if it does not bind to non-torsin polypeptides. Those which bind selectively to torsin would be chosen for use in methods which could include, but should not be limited to, the analysis of altered torsin expression in tissue containing torsin.

The torsin proteins of the present invention can be used in a variety of procedures and methods, such as for the generation of antibodies, for use in identifying pharmaceutical compositions, and for studying DNA/protein interaction.

The torsin peptide of the present invention can be used to produce antibodies or hybridomas. One skilled in the art will recognize that if an antibody is desired, such a peptide would be generated as described herein and used as an immunogen.

The antibodies of the present invention include monoclonal and polyclonal antibodies, as well as fragments of these antibodies. The invention further includes single chain antibodies. Antibody fragments which contain the idiotype of the molecule can be generated by known techniques. For example, such fragments include but are not limited to: the $F(ab')_2$ fragment; the Fab' fragments, Fab fragments, and Fv fragments.

Of special interest to the present invention are antibodies to torsin which are produced in humans, or are "humanized" (i.e. non-immunogenic in a human) by recombinant or other technology. Humanized antibodies can be produced, for example by replacing an immunogenic portion of an antibody with a corresponding, but non-immunogenic portion (i.e. chimeric antibodies) (Robinson, R.R. *et al.*, International Patent Publication PCT/US86/02269; Akira, K. *et al.*, European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison, S.L. *et al.*, European Patent Application 173,494; Neuberger, M.S. et al., PCT Application WO 86/01533; Cabilly, S. *et al.*, European Patent Application 125,023; Better, M. *et al.*, *Science 240*:1041-1043 (1988); Liu, A.Y. *et al.*, *Proc. Natl. Acad. Sci. USA 84*:3439-3443 (1987); Liu, A.Y. *et al.*, *J. Immunol. 139*:3521-3526 (1987); Sun, L.K. *et al.*, *Proc. Natl. Acad. Sci. USA 84*:214-218 (1987); Nishimura, Y. *et al.*, *Canc. Res. 47*:999-1005 (1987); Wood, C.R. *et al.*, *Nature 314*:446-449 (1985)); Shaw *et al.*, *J. Natl.Cancer Inst. 80*:1553-1559 (1988). General reviews of "humanized" chimeric antibodies are provided by Morrison, S.L. (*Science, 229*:1202-1207 (1985)) and by Oi, V.T. *et al.*, *BioTechniques 4*:214 (1986)). Suitable "humanized" antibodies can be alternatively produced by CDR or CEA substitution (Jones, P.T. *et al.*, *Nature 321*:552-525 (1986); Verhoeyan *et al.*, *Science 239*:1534 (1988); Beidler, C.B. *et al.*, *J. Immunol. 141*:4053-4060 (1988)).

In another embodiment, the present invention relates to a hybridoma which produces the above-described monoclonal antibody. A hybridoma is an

immortalized cell line which is capable of secreting a specific monoclonal antibody.

In general, techniques for preparing monoclonal antibodies and hybridomas are well known in the art (Campbell, "*Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology*," Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth *et al., J. Immunol. Methods 35*:1-21 (1980)).

Any animal (mouse, rabbit, and the like) which is known to produce antibodies can be immunized with the selected polypeptide. Methods for immunization are well known in the art. Such methods include subcutaneous or interperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of polypeptide used for immunization will vary based on the animal which is immunized, the antigenicity of the polypeptide and the site of injection.

The polypeptide can be modified or administered in an adjuvant in order to increase the peptide antigenicity. Methods of increasing the antigenicity of a polypeptide are well known in the art. Such procedures include coupling the antigen with a heterologous protein (such as globulin or β-galactosidase) or through the inclusion of an adjuvant during immunization.

For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells.

Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz *et al., Exp. Cell Res. 175*:109-124 (1988)).

Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell, *Monoclonal*

Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology, supra (1984)).

For polyclonal antibodies, antibody containing antisera is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures.

In another embodiment of the present invention, the above-described antibodies are detectably labeled. Antibodies can be detectably labeled through the use of radioisotopes, affinity labels (such as biotin, avidin, and the like), enzymatic labels (such as horse radish peroxidase, alkaline phosphatase, and the like) fluorescent labels (such as FITC or rhodamine, and the like), paramagnetic atoms, and the like. Procedures for accomplishing such labeling are well-known in the art, for example, see (Sternberger *et al., J. Histochem. Cytochem. 18*:315 (1970); Bayer *et al., Meth. Enzym. 62*:308 (1979); Engval *et al., Immunol. 109*:129 (1972); Goding, *J. Immunol. Meth. 13*:215 (1976)). The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues which express a specific peptide.

In another embodiment of the present invention the above-described antibodies are immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir *et al.*, "*Handbook of Experimental Immunology*" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby *et al., Meth. Enzym. 34* Academic Press, N.Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as in immunochromatography.

Furthermore, one skilled in the art can readily adapt currently available procedures, as well as the techniques, methods and kits disclosed above with regard to antibodies, to generate peptides capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides, for

example see Hurby *et al.*, "Application of Synthetic Peptides: Antisense Peptides", In *Synthetic Peptides, A User's Guide*, W.H. Freeman, NY, pp. 289-307 (1992), and Kaspczak *et al.*, *Biochemistry 28*:9230-8 (1989).

Anti-peptide peptides can be generated in one of two fashions. First, the anti-peptide peptides can be generated by replacing the basic amino acid residues found in the torsin peptide sequence with acidic residues, while maintaining hydrophobic and uncharged polar groups. For example, lysine, arginine, and/or histidine residues are replaced with aspartic acid or glutamic acid and glutamic acid residues are replaced by lysine, arginine or histidine.

VIII. A Method of Detecting a Torsin Polypeptide or Antibody in a Sample

In another embodiment, the present invention relates to a method of detecting a torsin polypeptide in a sample, comprising: a) contacting the sample with an above-described antibody (or protein), under conditions such that immunocomplexes form, and b) detecting the presence of the antibody bound to the polypeptide. In detail, the methods comprise incubating a test sample with one or more of the antibodies of the present invention and assaying whether the antibody binds to the test sample. Altered levels of torsin in a sample as compared to normal levels can indicate a specific disease.

In a further embodiment, the present invention relates to a method of detecting a torsin antibody in a sample, comprising: a) contacting the sample with an above-described torsin protein, under conditions such that immunocomplexes form, and b) detecting the presence of the protein bound to the antibody or antibody bound to the protein. In detail, the methods comprise incubating a test sample with one or more of the proteins of the present invention and assaying whether the antibody binds to the test sample.

Conditions for incubating an antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the antibody used in the assay. One skilled in the art will recognize that any one of the commonly available immunological

assay formats (such as radioimmunoassays, enzyme-linked immunosorbent assays, diffusion based Ouchterlony, or rocket immunofluorescent assays) can readily be adapted to employ the antibodies of the present invention. Examples of such assays can be found in Chard, *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock *et al.*, *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The immunological assay test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is capable with the system utilized.

IX. A Diagnostic Kit Comprising Torsin Protein or Antibody

In another embodiment of the present invention, a kit is provided which contains all the necessary reagents to carry out the previously described methods of detection.

The kit can comprise: i) a first container means containing an above-described antibody, and ii) second container means containing a conjugate comprising a binding partner of the antibody and a label.

The kit can comprise: i) a first container means containing an above-described protein, and preferably, ii) second container means containing a conjugate comprising a binding partner of the protein and a label. More

specifically, a diagnostic kit comprises torsin protein as described above, to detect antibodies in the serum of potentially infected animals or humans.

In another preferred embodiment, the kit further comprises one or more other containers comprising one or more of the following: wash reagents and reagents capable of detecting the presence of bound antibodies. Examples of detection reagents include, but are not limited to, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the chromophoric, enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. The compartmentalized kit can be as described above for nucleic acid probe kits.

One skilled in the art will readily recognize that the antibodies described in the present invention can readily be incorporated into one of the established kit formats which are well known in the art.

X. Diagnostic Screening and Treatment

It is to be understood that although the following discussion is specifically directed to human patients, the teachings are also applicable to any animal that expresses torsin.

The diagnostic and screening methods of the invention are especially useful for a patient suspected of being at risk for developing a disease associated with an altered expression level of torsin based on family history, or a patient in which it is desired to diagnose a torsin-related disease.

Preferably, DNA diagnosis is used as a means of differential diagnosis of various forms of dystonia. This information is then used in genetic counseling and in classifying patients with respect to individualized therapeutic strategies.

According to the invention, presymptomatic screening of an individual in need of such screening is now possible using DNA encoding the torsin protein of the invention. The screening method of the invention allows a presymptomatic diagnosis, including prenatal diagnosis, of the presence of a missing or aberrant

torsin gene in individuals, and thus an opinion concerning the likelihood that such individual would develop or has developed a torsin-associated disease. This is especially valuable for the identification of carriers of altered or missing torsin genes, for example, from individuals with a family history of a torsin-associated disease. Early diagnosis is also desired to maximize appropriate timely intervention.

Identification of gene carriers prior to onset of symptoms allows evaluation of genetic and environmental factors that trigger onset of symptoms. Modifying genetic factors could include polymorphic variations in torsin (specifically, torsinA) or mutations in related or associated proteins; environmental factors include sensory overload to the part of body subserved by susceptible neurons, such as that caused by overuse or trauma (Gasser, T., *et al.*, *Mov Disord 11:*163-166 (1996)); high body temperature; or exposure to toxic agents.

In one preferred embodiment of the method of screening, a tissue sample would be taken from such individual, and screened for (1) the presence of the "normal" torsin gene; (2) the presence of torsin mRNA and/or (3) the presence of torsin protein. The normal human gene can be characterized based upon, for example, detection of restriction digestion patterns in "normal" versus the patient's DNA, including RFLP analysis, using DNA probes prepared against the torsin sequence (or a functional fragment thereof) taught in the invention. Similarly, torsin mRNA can be characterized and compared to normal torsin mRNA (a) levels and/or (b) size as found in a human population not at risk of developing torsin-associated disease using similar probes. Lastly, torsin protein can be (a) detected and/or (b) quantitated using a biological assay for torsin activity or using an immunological assay and torsin antibodies. When assaying torsin protein, the immunological assay is preferred for its speed. An (1) aberrant torsin DNA size pattern, and/or (2) aberrant torsin mRNA sizes or levels and/or (3) aberrant torsin protein levels would indicate that the patient is at risk for developing a torsin-associated disease.

More specifically, a method of diagnosing the presence or predisposition to develop torsion dystonia in a patient is provided herein. The method comprises:

a) taking a sample from the patient;

b) evaluating the characteristics of torsinA nucleic acid in the sample, wherein the evaluation comprises detecting the GAGGAG region (SEQ ID NO:5 at nucleotide positions 946-951) in the sample; and

c) diagnosing the presence or predisposition to develop torsion dystonia in a patient wherein the absence of a GAG from the GAGGAG region indicates the presence or predisposition to develop torsion dystonia.

The screening and diagnostic methods of the invention do not require that the entire torsin DNA coding sequence be used for the probe. Rather, it is only necessary to use a fragment or length of nucleic acid that is sufficient to detect the presence of the torsin gene in a DNA preparation from a normal or affected individual, the absence of such gene, or an altered physical property of such gene (such as a change in electrophoretic migration pattern).

Prenatal diagnosis can be performed when desired, using any known method to obtain fetal cells, including amniocentesis, chorionic villous sampling (CVS), and fetoscopy. Prenatal chromosome analysis can be used to determine if the portion of the chromosome possessing the normal torsin gene is present in a heterozygous state.

In the method of treating a torsin-associated disease in a patient in need of such treatment, functional torsin DNA can be provided to the cells of such patient in a manner and amount that permits the expression of the torsin protein provided by such gene, for a time and in a quantity sufficient to treat such patient. Many vector systems are known in the art to provide such delivery to human patients in need of a gene or protein missing from the cell. For example, retrovirus systems can be used, especially modified retrovirus systems and

especially herpes simplex virus systems. Such methods are provided for, in, for example, the teachings of Breakefield, X.A. *et al., The New Biologist 3*:203-218 (1991); Huang, Q. *et al., Experimental Neurology 115*:303-316 (1992), WO93/03743 and WO90/09441. Delivery of a DNA sequence encoding a functional torsin protein will effectively replace the missing or mutated torsin gene of the invention.

In another embodiment of this invention, the torsin gene is expressed as a recombinant gene in a cell, so that the cells can be transplanted into a mammal, preferably a human in need of gene therapy. To provide gene therapy to an individual, a genetic sequence which encodes for all or part of the torsin gene is inserted into a vector and introduced into a host cell. Examples of diseases that can be suitable for gene therapy include, but are not limited to, neurodegenerative diseases or disorders, primary dystonia (preferably, generalized dystonia and torsion dystonia).

Gene therapy methods which can be used to transfer the torsin coding sequence of the invention to a patient are set forth in Chatterjee and Wong, *Current Topics in Microbiol. and Immuno., 218:* 61-73 (1996); Zhang, *J. Mol. Med. 74:*191-204 (1996); Schmidt-Wolf and Schmidt-Wolf, *J. of Hematotherapy 4:*551-561 (1995); Shaughnessy *et al., Seminars in Oncology 23(1):* 159-171 (1996); and Dunbar *Annu. Rev. Med. 47:*11-20 (1996).

Examples of vectors that may be used in gene therapy include, but are not limited to, defective retroviral, adenoviral, or other viral vectors (Mulligan, R.C., *Science 260*:926-932 (1993)). The means by which the vector carrying the gene can be introduced into the cell include but is not limited to, microinjection, electroporation, transduction, or transfection using DEAE-Dextran, lipofection, calcium phosphate or other procedures known to one skilled in the art (*Molecular Cloning, A Laboratory Manual*, Sambrook *et al.*, eds., Cold Spring Harbor Press, Plainview, New York (1989)).

The ability of antagonists and agonists of torsin to interfere or enhance the activity of torsin can be evaluated with cells containing torsin. An assay for

torsin activity in cells can be used to determine the functionality of the torsin protein in the presence of an agent which may act as antagonist or agonist, and thus, agents that interfere or enhance the activity of torsin are identified.

The agents screened in the assays can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. These agents can be selected and screened 1) at random, 2) by a rational selection or 3) by design using for example, protein or ligand modeling techniques (preferably, computer modeling).

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to or stimulate/block the activity of the torsin protein.

Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the torsin protein.

In one embodiment, the present invention relates to a method of screening for an antagonist or agonist which stimulates or blocks the activity of torsin comprising:

(a) incubating a cell expressing torsin with an agent to be tested; and

(b) assaying the cell for the activity of the torsin protein by measuring the agents effect on ATP binding of torsin.

Any cell may be used in the above assay so long as it expresses a functional form of torsin and the torsin activity can be measured. The preferred expression cells are eukaryotic cells or organisms. Such cells can be modified to contain DNA sequences encoding torsin using routine procedures known in the art. Alternatively, one skilled in the art can introduce mRNA encoding the torsin protein directly into the cell.

In another embodiment, the present invention relates to a screen for pharmaceuticals (ie. drugs) which can counteract the expression of a mutant torsin protein. Preferably, a neuronal culture is used for the overexpression of the mutant form of torsin A using the vector technology described herein. Changes

in neuronal morphology and protein distribution is assessed and a means of quantification is used. This bioassay is then used as a screen for drugs which can ameliorate the phenotype.

Using torsin ligands (including antagonists and agonists as described above) the present invention further provides a method for modulating the activity of the torsin protein in a cell. In general, agents (antagonists and agonists) which have been identified to block or stimulate the activity of torsin can be formulated so that the agent can be contacted with a cell expressing a torsin protein *in vivo*. The contacting of such a cell with such an agent results in the *in vivo* modulation of the activity of the torsin proteins. So long as a formulation barrier or toxicity barrier does not exist, agents identified in the assays described above will be effective for *in vivo* use.

In another embodiment, the present invention relates to a method of administering torsin or a torsin ligand (including torsin antagonists and agonists) to an animal (preferably, a mammal (specifically, a human)) in an amount sufficient to effect an altered level of torsin in the animal. The administered torsin or torsin ligand could specifically effect torsin associated functions. Further, since torsin is expressed in brain tissue, administration of torsin or torsin ligand could be used to alter torsin levels in the brain.

One skilled in the art will appreciate that the amounts to be administered for any particular treatment protocol can readily be determined. The dosage should not be so large as to cause adverse side effects, such as unwanted cross-reactions, anaphylactic reactions, and the like. Generally, the dosage will vary with the age, condition, sex and extent of disease in the patient, counter indications, if any, and other such variables, to be adjusted by the individual physician. Dosage can vary from .001 mg/kg to 50 mg/kg of torsin or torsin ligand, in one or more administrations daily, for one or several days. Torsin or torsin ligand can be administered parenterally by injection or by gradual perfusion over time. It can be administered intravenously, intraperitoneally, intramuscularly, or subcutaneously.

Preparations for parenteral administration include sterile or aqueous or non-aqueous solutions, suspensions, and emulsions. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Aqueous carriers include water, alcoholic/aqueous solutions, emulsions or suspensions, including saline and buffered media. Parenteral vehicles include sodium chloride solution, Ringer's dextrose and sodium chloride, lactated Ringer's, or fixed oils. Intravenous vehicles include fluid and nutrient replenishers, electrolyte replenishers, such as those based on Ringer's dextrose, and the like. Preservatives and other additives can also be present, such as, for example, antimicrobials, antioxidants, chelating agents, inert gases and the like. See, generally, *Remington's Pharmaceutical Science*, 16th Ed., Mack Eds. (1980).

In another embodiment, the present invention relates to a pharmaceutical composition comprising torsin or torsin ligand in an amount sufficient to alter torsin associated activity, and a pharmaceutically acceptable diluent, carrier, or excipient. Appropriate concentrations and dosage unit sizes can be readily determined by one skilled in the art as described above (See, for example, *Remington's Pharmaceutical Sciences* (16th ed., Osol, A., Ed., Mack, Easton PA (1980) and WO 91/19008).

XI. Transgenic Torsin "Knock-Out" Mice

Methods of Generating Transgenic Non-Human Animals

The non-human animals of the invention comprise any animal having a transgenic interruption or alteration of the endogenous gene(s) (knock-out animals) and/or into the genome of which has been introduced one or more transgenes that direct the expression of human torsin.

Such non-human animals include vertebrates such as rodents, non-human primates, sheep, dog, cow, amphibians, reptiles, etc. Preferred non-human animals are selected from non-human mammalian species of animals, most

preferably, animals from the rodent family including rats and mice, most preferably mice.

The transgenic animals of the invention are animals into which has been introduced by nonnatural means (i.e., by human manipulation), one or more genes that do not occur naturally in the animal, e.g., foreign genes, genetically engineered endogenous genes, etc. The nonnaturally introduced genes, known as transgenes, may be from the same or a different species as the animal but not naturally found in the animal in the configuration and/or at the chromosomal locus conferred by the transgene. Transgenes may comprise foreign DNA sequences, i.e., sequences not normally found in the genome of the host animal. Alternatively or additionally, transgenes may comprise endogenous DNA sequences that are abnormal in that they have been rearranged or mutated *in vitro* in order to alter the normal *in vivo* pattern of expression of the gene, or to alter or eliminate the biological activity of an endogenous gene product encoded by the gene. (Watson, J.D., *et al.*, in *Recombinant DNA*, 2d Ed., W.H. Freeman & Co., New York (1992), pages 255-272; Gordon, J.W., *Intl. Rev. Cytol. 115*:171-229 (1989); Jaenisch, R., *Science 240*:1468-1474 (1989); Rossant, J., *Neuron 2*:323-334 (1990)).

The transgenic non-human animals of the invention are produced by introducing transgenes into the germline of the non-human animal. Embryonic target cells at various developmental stages are used to introduce the transgenes of the invention. Different methods are used depending on the stage of development of the embryonic target cell(s).

1. Microinjection of zygotes is the preferred method for incorporating transgenes into animal genomes in the course of practicing the invention. A zygote, a fertilized ovum that has not undergone pronuclei fusion or subsequent cell division, is the preferred target cell for microinjection of transgenic DNA sequences. The murine male pronucleus reaches a size of approximately 20 micrometers in diameter, a feature which allows for the reproducible injection of 1-2 picoliters of a solution containing transgenic DNA sequences. The use of a

zygote for introduction of transgenes has the advantage that, in most cases, the injected transgenic DNA sequences will be incorporated into the host animal's genome before the first cell division (Brinster, *et al.*, *Proc. Natl. Acad. Sci. (USA) 82*:4438-4442 (1985)). As a consequence, all cells of the resultant transgenic animals (founder animals) stably carry an incorporated transgene at a particular genetic locus, referred to as a transgenic allele. The transgenic allele demonstrates Mendelian inheritance: half of the offspring resulting from the cross of a transgenic animal with a non-transgenic animal will inherit the transgenic allele, in accordance with Mendel's rules of random assortment.

2. Viral integration can also be used to introduce the transgenes of the invention into an animal. The developing embryos are cultured *in vitro* to the developmental stage known as a blastocyst. At this time, the blastomeres may be infected with appropriate retroviruses (Jaenich, R., *Proc. Natl. Sci. (USA) 73*:1260-1264 (1976)). Infection of the blastomeres is enhanced by enzymatic removal of the zona pellucida (Hogan, *et al.*, in *Manipulating the Mouse Embryo*, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1986)). Transgenes are introduced via viral vectors which are typically replication-defective but which remain competent for integration of viral-associated DNA sequences, including transgenic DNA sequences linked to such viral sequences, into the host animal's genome (Jahner, *et al.*, *Proc. Natl. Acad. Sci. (USA) 82*:6927-6931 (1985); Van der Putten, *et al.*, *Proc. Natl. Acad. Sci. (USA) 82*:6148-6152 (1985)). Transfection is easily and efficiently obtained by culture of blastomeres on a mono-layer of cells producing the transgene-containing viral vector (Van der Putten, *et al.*, *Proc. Natl. Acad. Sci. (USA) 82*:6148-6152 (1985); Stewart, *et al.*, *EMBO Journal 6*:383-388 (1987)). Alternatively, infection may be performed at a later stage, such as a blastocoele (Jahner, D., *et al.*, *Nature 298*:623-628 (1982)). In any event, most transgenic founder-animals produced by viral integration will be mosaics for the transgenic allele; that is, the transgene is incorporated into only a subset of all the cells that form the transgenic founder animal. Moreover, multiple viral integration events may occur in a single founder

animal, generating multiple transgenic alleles which will segregate in future generations of offspring. Introduction of transgenes into germline cells by this method is possible but probably occurs at a low frequency (Jahner, D., *et al.*, *Nature 298*:623-628 (1982)). However, once a transgene has been introduced into germline cells by this method, offspring may be produced in which the transgenic allele is present in all of the animal's cells, i.e., in both somatic and germline cells.

3. Embryonic stem (ES) cells can also serve as target cells for introduction of the transgenes of the invention into animals. ES cells are obtained from pre-implantation embryos that are cultured *in vitro* (Evans, M.J., *et al.*, *Nature 292*:154-156 (1981); Bradley, M. O., *et al.*, *Nature 309*:255-258 (1984); Gossler, *et al.*, *Proc. Natl. Acad. Sci. (USA) 83*:9065-9069 (1986); Robertson *et al.*, *Nature 322*:445-448 (1986); Robertson, E.J., in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, E.J., ed., IRL Press, Oxford (1987), pages 71-112). ES cells, which are commercially available (from, e.g., Genome Systems, Inc., St. Louis, MO), can be transformed with one or more transgenes by established methods (Lovell-Badge, R.H., in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, E.J., ed., IRL Press, Oxford (1987), pages 153-182). Transformed ES cells can be combined with an animal blastocyst, whereafter the ES cells colonize the embryo and contribute to the germline of the resulting animal, which is a chimera (composed of cells derived from two or more animals) (Jaenisch, R., *Science 240*:1468-1474 (1988); Bradley, A., in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, E.J., ed., IRL Press, Oxford (1987), pages 113-151). Again, once a transgene has been introduced into germline cells by this method, offspring may be produced in which the transgenic allele is present in all of the animal's cells, i.e., in both somatic and germline cells.

However it occurs, the initial introduction of a transgene is a Lamarckian (non-Mendelian) event. However, the transgenes of the invention may be stably integrated into germ line cells and transmitted to offspring of the transgenic

animal as Mendelian loci. Other transgenic techniques result in mosaic transgenic animals, in which some cells carry the transgenes and other cells do not. In mosaic transgenic animals in which germ line cells do not carry the transgenes, transmission of the transgenes to offspring does not occur. Nevertheless, mosaic transgenic animals are capable of demonstrating phenotypes associated with the transgenes.

Transgenes may be introduced into non-human animals in order to provide animal models for human diseases. Transgenes that result in such animal models include, e.g., transgenes that encode mutant gene products associated with an inborn error of metabolism in a human genetic disease and transgenes that encode a human factor required to confer susceptibility to a human pathogen (i.e., a bacterium, virus, or other pathogenic microorganism) (Leder *et al.*, U.S. Patent 5,175,383 (Dec. 29, 1992); Kindt *et al.*, U.S. Patent 5,183,949 (Feb. 2, 1993); Small *et al.*, *Cell 46*:13-18 (1986); Hooper *et al.*, *Nature 326*:292-295 (1987); Stacey *et al.*, *Nature 332*:131-136 (1988); Windle *et al.*, *Nature 343*:665-669 (1990); Katz *et al.*, *Cell 74*:1089-1100 (1993)). Transgenically introduced mutations comprise null ("knock-out") alleles in which a DNA sequence encoding a selectable and/or detectable marker is substituted for a genetic sequence normally endogenous to a non-human animal. Resultant transgenic non-human animals that are predisposed to a disease, or in which the transgene causes a disease, may be used to identify compositions that induce the disease and to evaluate the pathogenic potential of compositions known or suspected to induce the disease (Berns, A.J.M., U.S. Patent 5,174,986 (Dec. 29, 1992)), or to evaluate compositions which may be used to treat the disease or ameliorate the symptoms thereof (Scott *et al.*, WO 94/12627 (1994)).

Offspring that have inherited the transgenes of the invention are distinguished from littermates that have not inherited transgenes by analysis of genetic material from the offspring for the presence of biomolecules that comprise unique sequences corresponding to sequences of, or encoded by, the transgenes of the invention. For example, biological fluids that contain

polypeptides uniquely encoded by the selectable marker of the transgenes of the invention may be immunoassayed for the presence of the polypeptides. A more simple and reliable means of identifying transgenic offspring comprises obtaining a tissue sample from an extremity of an animal, e.g., a tail, and analyzing the sample for the presence of nucleic acid sequences corresponding to the DNA sequence of a unique portion or portions of the transgenes of the invention, such as the selectable marker thereof. The presence of such nucleic acid sequences may be determined by, e.g., hybridization ("Southern") analysis with DNA sequences corresponding to unique portions of the transgene, analysis of the products of PCR reactions using DNA sequences in a sample as substrates and oligonucleotides derived from the transgene's DNA sequence, etc.

XII. HSV-1 Amplicon Constructs

In another embodiment, the present invention relates to a recombinant DNA molecule comprising an HSV-1 amplicon and at least one above-described torsin nucleic acid molecule.

Several features make HSV-1 an ideal candidate for vector development: (i) HSV-1 is essentially pantropic and can infect both dividing and non-dividing cells, such as neurons and hepatocytes; (ii) the HSV-1 genome can remain in neurons for long periods with at least some transcriptional activity; (iii) the HSV-1 genome encodes more than 75 genes of which 38 are dispensable (non-essential) for virus replication in cell culture (Ward, P.L. and Roizman, B., *Trends Genet. 10*:267-274 (1994)). This offers the opportunity to replace large parts of the genome with foreign DNA, including one or more therapeutic genes of interest.

The technology to construct recombinant HSV-1 vectors was developed more than a decade ago (Mocarski, E.S., *et al., Cell 22*:243-255 (1980); Post, L.E. and Reizman, B., *Cell 25*:2227-2232 (1981); Roizman, B. and Jenkins, F.J., *Science 229*:1208-1214 (1985)). With the goal to create a prototype HSV-

1/HSV-2 recombinant vaccine, the HSV-1 genome was deleted in certain domains in order to eliminate some loci responsible for neurovirulence, such as the viral thymidine kinase gene, and to create space for the insertion of a DNA fragment encoding the herpes simplex virus type 2 (HSV-2) glycoproteins D, G, and I (Meignier, B., *et al., J. Inf. Dis. 158*:602-614 (1988)). Currently, recombinant herpes virus vectors are being evaluated in numerous protocols primarily for gene therapy of neurodegenerative diseases and brain tumors (Breakefield, X.O., *et al., Cancer Gene Therapeutics*, (1995), pp. 41-56; Glorioso, J.C., *et al.*, "Herpes simplex virus as a gene-delivery vector for the central nervous system," in *Viral vectors: Gene therapy and neuroscience applications*, Kaplitt, M.G. and Loewy, A.D., eds., Academic Press, New York, NY (1995), pp. 1-23).

The development of a second type of HSV-1 vector, the so-called HSV-1 "amplicon" vector (Spaete and Frenkel, 1982), was based on the characterization of naturally occurring defective HSV-1 genomes (Frenkel, N., *et al., J. Virol. 20*:527-531 (1976)). Amplicons carry three types of genetic elements: (i) prokaryotic sequences for propagation of plasmid DNA in bacteria, including an *E. coli* origin of DNA replication and an antibiotic resistance gene; (ii) sequences from HSV-1, including an *ori* and a *pac* signal to support replication and packaging into HSV-1 particles in mammalian cells in the presence of helper virus functions; and (iii) a transcription unit with one or more genes of interest (Ho, D.Y., *Meth. Cell. Biol. 43*:191-210 (1994); Kwong, A.D., and Frenkel, N., "Biology of herpes simplex virus (HSV) defective viruses and development of the amplicon system, in *Viral vectors: Gene therapy and neuroscience applications*, Kaplitt, M.G., and Loewy, A.D., eds., Academic Press, New York, NY (1995), pp. 25-42).

Replication of amplicon DNA in mammalian cells is mediated by interaction of the HSV-1 origins of DNA replication *(ori*$_s$ or *ori*$_L$*)* with proteins provided in *trans* by the helper virus. These include: (i) the products of the *UL5, 8* and *52* genes which form a complex that has helicase-primase activity; (ii) the

UL9 gene product, which binds directly to *ori;* (iii) a single stranded DNA binding protein (the product of *UL29),* which forms a complex with the products of *UL42,* a double stranded DNA binding protein, and the *UL30* gene product, which is a virus encoded DNA polymerase (Ward, P.L. and Roizman, B., *Trends Genet. 10*:267-274 (1994)). The ori_L sequence (144 bp) is an A + T rich palindrome which is unstable in bacteria due to its dyad symmetry (Weller, S.K., *et al., Mol. Cell Biol. 5*:930-942 (1985)) and thus has not proven useful in the generation of amplicon vectors (Kwong, A.D., and Frenkel, N., "Biology of herpes simplex virus (HSV) defective viruses and development of the amplicon system, in *Viral vectors: Gene therapy and neuroscience applications*, Kaplitt, M.G., and Loewy, A.D., eds., Academic Press, New York, NY (1995), pp. 25-42). The ori_s sequence (90 bp), in contrast, with a shorter A + T rich sequence and imperfect palindrome, has proven more stable in bacteria (Stow, N.D. and McMonagle, E.C., *Virology 130*:427-438 (1983)) and is typically incorporated into amplicons. Ori_s sequences from HSV-1 (Ho, 1994) and HSV-2 (Kaplitt, M.G., *et al., Mol. Cell. Neurosci. 2*:320-330 (1991)) have both been used, although the HSV-1 one is the most typical. HSV-1 ori_s is located between the promoters for the immediate-early *(IE) 3* and *4/5* genes (Fig. 2A). These promoters contain TAATGARAT sequences which respond to the virion tegument protein, VP16, as well as SP1 enhancer elements (Stern, S., *Nature 341*:624-630 (1989)), but they also increase the efficiency of DNA replication (Wong, S.W. and Schaffer, P.A., *J. Virol. 65*:2601-2611 (1991)). Thus larger fragments (0.5 - 1 kb) bearing ori_s can be used both for efficient amplicon replication in mammalian cells and to direct the expression of the transgenes from the *IE3* and/or *IE4/5* promoters. However, to reduce the non-specific effects of these viral regulatory elements on cell-specific or inducible promoters, several groups have used minimal ori_s elements (237-295 bp) without compromising the efficient generation of amplicon vectors (Kaplitt, M.G., *et al., Mol. Cell. Neurosci. 2*:320-330 (1991); Ho, D.Y., *Meth. Cell. Biol. 43*:191-210 (1994); Lu, B. and Federoff, H.J., *Hum. Gene Ther. 6*:419-428 (1995)).

Replication of amplicon DNA in cells co-transduced with HSV-1 helper virus proceeds by a rolling-circle mechanism creating a linear concatemer of amplicon DNA sequences. For packaging into virion capsids concatemeric genomes are cleaved between pairs of a repeat sequences after filling of the capsid, which holds about 152 kb of DNA (Deiss, L.P., and Frenkel, N., *J. Virol. 59*:605-618 (1986); Roizman, B. and Sears, A.E., "Herpes simplex viruses and their replication," in *Virology*, 3rd. edition, Fields, B.N., *et al.*, eds., Lippincott-Raven, Philadelphia, PA (1996), pp. 2231-2295). The *pac* sequences within the a repeat define the cleavage point and consist of alternating repeat and unique sequences of 250-500 bp, depending on the virus strain, in the following configuration: direct repeat (*DR*) *1* (20 bp) - unique sequence (*U*) *b* (65 bp) - *DR2* (12 bp x 19-23) - *DR4* (37 bp x 2-3) - *Uc* (58 bp) - *DR1* (Davison, A.J., and Wilkie, N.M., *J. Gen. Virol. 55*:315-331 (1981)). The repeat nature of these sequences may contribute to their instability in bacteria, and in addition, they contain elements which can serve as recombinational hot spots in the context of the HSV-1 infection of mammalian cells (Umene, K., *J. Virol. 67*:5685-5691 (1993)). Also, the promoter of the γ34.5 gene, which is located in the a repeat, can potentially influence transgene expression mediated by amplicon vectors.

Current vectors contain one to two transgene cassettes (Kaplitt, M.G., *et al., Mol. Cell. Neurosci. 2*:320-330 (1991); Ho, D.Y., *Meth. Cell. Biol. 43*:191-210 (1994); Linnik, M.D., *et al., Stroke 26*:1670-1674 (1995); Lawrence, M.S., *et al., Blood Flow Metab. 16*:181-185 (1996); New, K. and Rabkin, S., *Mol. Brain Res. 37*:317-323 (1996); Pechan, P.A., *et al., Hum. Gene Ther. 7*:2003-2013 (1996)), but even within the 15 kb limit, three or more genes could be included, depending on the size of the transgenes.

A number of different promoter elements have been used to regulate transgene expression in mammalian cells, including HSV-1 *IE* promoters and cell-specific promoters (Smith, R.L., *et al., J. Virol. 69*:4593-4599 (1995)). Since the *IE* promoters are induced by the tegument protein, VP16, which is carried into the nucleus by the virion, they tend to give robust expression in the first few days

after infection and then decrease dramatically as VP16 is degraded by the cells. However, *IE* promoters can be re-activated by superinfection with HSV-1 (Starr, P.A., *et al., Gene Ther. 3*:615-623 (1996)). Other viral promoters, including hCMV *IE1* and SV40 *T*, also direct strong but transient transgene expression in most cells (Ho, D.Y., *et al., Proc. Acad. Natl. Sci. USA 90*:3655-3659 (1993); Pechan, P.A., *et al., Hum. Gene Ther. 7*:2003-2013 (1996)). Several groups have utilized cell specific promoters in the context of amplicon vectors, including those for preproenkephalin (Kaplitt, M.G., *et al., Proc. Natl. Acad. Sci. USA 91*:8979-8983 (1994)), neurofilament light and heavy gene (Fraefel, C., *et al., 21st Intl. Herpesvirus Workshop*, DeKalb IL (1996)), tyrosine hydroxylase (TH; Oh, Y.J., *et al., Mol. Brain Res. 35*:227-236 (1996); Jin, B.K., *et al., Hum. Gene Ther. 7*:2-15-2024 (1996); Fraefel, C., *et al., 21st Intl. Herpesvirus Workshop*, DeKalb IL (1996)), neuron specific enolase, sodium channel, albumin, and αl-antitrypsin (C. Fraefel, unpublished material). Some of these promoters appear to retain their cell specificity in the context of amplicon sequences, although levels of expression tend to be lower than with viral promoters. Moreover, the extent of specificity is difficult to assess given the altered transcriptional regulation in neural cells in culture versus *in vivo* and the difficulty in identifying neural cell types in *vivo*. Only two reports have demonstrated inducible expression mediated by amplicon vectors. Using a minimal *ori*$_s$ sequence (234 bp), Lu and Federoff (Lu, B. and Federoff, H.J., *Hum. Gene Ther. 6*:419-428 (1995)) were able to achieve up to 50-fold dexamethasone induction of *lacZ* expression in primary rat hepatocytes using five copies of a tandemly repeated rat tyrosine aminotransferase (TAT) glucocorticoid responsive element GRE; Jantzen, H.M., *et al., Cell 49*:29-38 (1987)).

Cell lines have been developed which allow the helper virus-free packaging of certain other vectors, such as retrovirus vectors (Mann, R., *et al., Cell 33*:153-159 (1983)). A similar approach for the generation of HSV-1 amplicon vectors, however, would presumably require both the expression of at least the 38 essential genes of HSV-1 (Ward, P.L. and Roizman, B., *Trends*

Genet. 10:267-274 (1994); Roizman, B. and Sears, A.E., "Herpes simplex viruses and their replication," in *Virology*, 3rd. edition, Fields, B.N., *et al.*, eds., Lippincott-Raven, Philadelphia, PA (1996), pp. 2231-2295) and replication of the HSV-1 genome which is essential for the expression of some late genes (Mavromara-Nazos, P. and Roizman, B., *Virology 161*:593-598 (1989)). To circumvent these problems, a helper virus-free packaging system has been developed which utilized transient cotransfection of amplicon DNA with a set of five cosmids that overlap and represent the entire HSV-1 genome but which were mutated to inactivate the *pac* signals (Fraefel, C., *et al., J. Virol. 70*:7190-7197 (1996); Cunningham, C., and Davison, A.J., *Virology 197*:116-124 (1993)) demonstrated that after transfection of cells, a HSV-1 cosmid set could form a complete replication-competent virus genome, via homologous recombination between the overlapping sequences, and produce infectious virus particles. By deleting the *pac* signals, however, these virus genomes are not packagable, while still providing all the helper-functions required for the replication and packaging of the cotransfected amplicon DNA (Fraefel, C., *et al., J. Virol. 70*:7190-7197 (1996)). The resulting vector stocks are free of detectable helper virus and have titers of 106 to 107 infectious vector particles per mililiter of culture medium. Moreover, in the absence of helper virus, these vector stocks can efficiently transduce many different cell types, including neural cells and hepatocytes in culture and *in vivo*, while causing minimal cytopathic effects (Fraefel, C., *et al., J. Virol. 70*:7190-7197 (1996); Fraefel, C., *et al., 21st Intl. Herpesvirus Workshop*, DeKalb IL (1996); Johnston, K.M., *et al., Hum. Gene Ther. 8*:359-370 (1997)).

The basic structure of amplicon vectors has remained relatively unchanged and includes the HSV-1 *ori*$_s$ and *pac* elements responsible for replication and packaging of constructs into HSV-1 virions. New variations have evolved in response to the demand for stable and cell specific expression. These include the use of: (i) different promoters, (ii) multiple transgenes, and (iii) elements from other virus vectors.

In the central nervous system (CNS), viral promoters such as the hCMV *IE1* promoter and the HSV-1 *IE4/5* promoter, support strong expression of amplicon delivered transgenes (Ho, D.Y., *et al., Proc. Acad. Natl. Sci. USA 90*:3655-3659 (1993); Smith, R.L., *et al., J. Virol. 69*:4593-4599 (1995)). In general, however, the time span of expression is relatively short-lived, due to poorly understood promoter inactivation. Cell type specific expression in the CNS has been attempted using several promoters, and specificity in the context of an amplicon has been reported using the neuronal preproenkephalin (PPE) promoter (Kaplitt, M.G., *et al., Proc. Natl. Acad. Sci. USA 91*:8979-8983 (1994)) and the TH promoter (Jin, B.K., *et al., Hum. Gene Ther. 7*:2-15-2024 (1996)). In these reports, cell specificity is due to the incorporation of large 5' regulatory sequences upstream of the minimal promoter elements of these mammalian genes. After amplicon vector delivery to the rat brain, the 2.7 kb PPE promoter and regulatory regions supported the expression of an *E. coli lacZ* marker gene for 2 months in cells morphologically resembling neurons. In transgenic mice, the 9.0 kb TH promoter/regulatory sequence efficiently directs cell and region specific expression of a lacZ marker gene (Min, N., *et al., Mol. Brain Res. 27*:281-289 (1994)). The importance of this large upstream region was confirmed by comparative analysis of two amplicons with *lacZ* marker genes driven by either the minimal HSV-1 *IE4/5* promoter or the 9.0 kb TH promoter (Jin, B.K., *et al., Hum. Gene Ther. 7*:2-15-2024 (1996)). Both amplicons supported transient synthesis of β-galactosidase at the site of inoculation. However, the TH amplicon continued to express the lacZ gene at apparently similar levels and in the same number of cells for up to ten weeks post-inoculation. As confirmed by double labeling experiments, expression was anatomically restricted to neurons in the substantia nigra (SN) and locus coeruleus (LC), where endogenous catecholamines are synthesized. These reports strongly support other studies indicating that the genetic elements which contribute to cell specificity and long-term expression are contained in the 5' regions upstream of the minimal promoter elements (Jin, B.K., *et al., Hum. Gene Ther. 7*:2-15-2024 (1996)). The ability of

HSV-1 amplicons to carry these large regulatory sequences makes this a valuable system for cell-specific gene expression in the CNS.

HSV-1 amplicons have been designed that transfer multiple transgenes using bicistronic genes or multiple expression cassettes. The picornavinis 5' ribosome binding region has been used as an internal entry site, thus linking the expression of two coding regions to a single promoter. This approach has been employed to "tag" the expression of a therapeutic gene with a marker gene. For example, amplicons overexpressing the rat brain glucose transporter (GLUT-1), linked by the internal ribosome entry site *(IRES)* to the *E. coli lacZ* gene, have been used in several experimental models (Ho, D.Y., *et al., J. Neurochem. 65*:842-850 (1995); Dash, R., *et al., Exp. Neurol. 137*:43-48 (1996)). These authors demonstrated protection of neuronal death against hypoglycemia, glutamate, and 3-nitroproprionic acid. The GLUT-1/*lacZ* bicistronic construct allowed the authors to conclude that there was an inverse correlation between the expression of the delivered transgenes and the degree of hippocampal neuron loss in a kainate induced seizure model (Lawrence, M.S., *et al., Proc. Natl. Acad. Sci. USA 92*:7247-7251 (1995); Lawrence, M.S., *et al., Blood Flow Metab. 16*:181-185 (1996)). New and Rabkin (1996) incorporated two independent marker genes, *lacZ* and alkaline phosphatase (AP), both under control of human cytomegalovirus (hCMV) *IE1* promoters. In a variety of cell lines and primary neurons in culture, as well as neurons *in vivo,* this vector supported the expression of both transgene products in about 40% of labeled cells. In the remaining labeled cells, only AP or β-galactosidase was detected at about the same rate (~30%).

The so-called "piggyback" system has been designed to improve the amplicon packaging efficiency (Pechan, P.A., *et al., Hum. Gene Ther. 7*:2003-2013 (1996)). In traditional packaging systems, cells infected with helper virus alone generate more helper virus, and those transfected with amplicon plasmid can generate vector particles only if infected with helper virus, and consequently, they also produce helper virus. This discrepancy favors the production of helper

virus and ensures that the ratios of amplicon vector to helper virus ratios remains low (usually <1). The "piggyback" amplicon encodes an HSV-1 gene essential for replication (*IE3*) which can complement a replication-incompetent helper virus deleted in *IE3*. In this system, packaging can be performed on any susceptible cell line as it does not need to complement the replication-defective helper virus. Packaging occurs only in cells in which both the helper virus and the amplicon are present, thus eliminating helper virus propagated independently of the amplicon. This system supports high amplicon titers of 6×10^7 transducing units per ml (t.u./ml), requires fewer passages to generate high titer vector stocks, and has apparent ratios of amplicon vector to helper virus of 3 to 5.

The newest generation of amplicon vectors incorporates both multiple transgenes and multiple genetic elements taken from other virus-based vectors. These "hybrid" amplicons are packaged in HSV-1 virions, and therefore retain the advantages of this virus for gene delivery to the CNS (see below), but include elements which are predicted to maintain the vector in a stable state that could support long-term gene expression in transduced cells. Wang and Vos (1996) constructed an amplicon vector that includes HSV-1 elements for replication and packaging, as well as the Epstein-Barr virus (EBV) nuclear antigen 1 (EBNA-1) gene and *oriP*, and the hygromycin resistance gene (HygR). EBV elements were included to support the autonomous replication of the vector genome and to maintain it as a stable episome in the host cell nucleus. After transfection of the amplicon into cells, stable colonies were isolated by hygromycin selection, and packaged vectors were generated by superinfecting these colonies with replication-incompetent helper HSV-1. Because all cells contain the hybrid amplicon, packaging was comparably efficient to the piggyback system (4×10^6 t.u./ml), and resulted in high ratios of amplicon vector to helper virus. This vector was successfully used to infect a number of human cell lines in culture, and two human tumor lines, the hepatoma line HepG2 and the glioma line T98G *in vivo*. Expression of a *lacZ* transgene was noted for at least two weeks after delivery (Wang, S. and Vos, J., *J. Virol. 70*:8422-8430 (1996)).

The second hybrid amplicon system incorporates the adeno-associated virus (AAV) inverted terminal repeats *(ITRs)*, and the AAV *rep* gene in addition to the HSV-1 replication and packaging elements (Fig. 2B). The AAV Rep isozymes have several functions, including the recognition of the *ITRs* for replication of the virus genome and subsequent integration in a site-specific manner into human chromosome 19q13 (Samulski, R.J., *et al., EMBO J. 10*:3941-3950 (1991); Berns, K.I., "*Paravoviridae*: the viruses and thier replication," in *Fields Virology*, 3rd edition, Fields, B.N., *et al.*, eds., Lippincott-Raven, Philadelphia, PA (1996), pp. 2173-2197). It has been postulated that these functions would allow a transgene flanked by the *ITRs* to be amplified from the hybrid vector and then to remain in the transduced cell as a stable provirus, integrated in a directed manner into this non-essential locus (Johnston, K.M., *et al., Hum. Gene Ther. 8*:359-370 (1997)). These events should be able to occur in both dividing and post-mitotic cells (Berns, K.I., "*Paravoviridae*: the viruses and thier replication," in *Fields Virology*, 3rd edition, Fields, B.N., *et al.*, eds., Lippincott-Raven, Philadelphia, PA (1996), pp. 2173-2197). These hybrid amplicons have been packaged by using both HSV-1 helper virus and helper virus-free systems (Johnston, K.M., *et al., Hum. Gene Ther. 8*:359-370 (1997)). The replicative and packaging functions of the AAV *ITRs* in the context of the hybrid amplicon was confirmed by co-transfecting the hybrid amplicon with a plasmid that carries AAV helper functions in the presence of either adenovirus or HSV-1 helper virus. Under these conditions, an *ITR*-flanked marker *lacZ* transgene was efficiently excised from the HSV/AAV amplicon, replicated, and packaged as a recombinant AAV. The HSV/AAV amplicon vector, in comparison to a conventional HSV-1 amplicon, supported extended transgene expression in dividing human U87 glioma cells. At fifteen days post-infection, the HSV/AAV amplicon packaged with helper virus produced about 100-fold more β-galactosidase-positive cells than conventional amplicons. Interestingly, although the helper virus-free packaged HSV/AAV amplicon vector had no demonstrable toxicity, gene expression was only 10-fold greater than with the

conventional amplicon, suggesting that the helper virus may augment HSV/AAV-mediated gene expression. At fifteen days post-transduction, the hybrid amplicon-transduced cells contained significantly more *lacZ* transgene DNA than]cells transduced with conventional amplicons, as determined by PCR analysis. The two hybrid amplicon systems discussed here are designed to extend transgene expression by adding genetic elements from other vector systems that support stable retention of the vector genome in target cells. The packaging of these hybrid amplicon vectors as defective virions ensures efficient and safe delivery of transgenes to the nucleus of host cells.

a. Amplicon Vectors for Gene Transfer into Neurons

In another embodiment, the present invention relates to the use of the above-described amplicon vectors for transfer of a torsin nucleic acid molecule into neurons.

HSV-1 has several biological properties that facilitate its use as a gene transfer vector into the CNS. These include: (i) a large transgene capacity (theoretically up to 150 kb), (ii) tropism for the CNS *in vivo*, (iii) nuclear localization in dividing as well as nondividing cells, (iv) a large host cell range in tissue culture, (v) the availability of a panel of neuroattenuated and replication incompetent mutants, and (vi) the possibility to produce relatively high virus titers.

Another important property of the HSV-1 derived vector systems for the CNS is the ability of these virions to be transported retrogradely along axons. After fusion with the cell membrane, the virus capsid and associated tegument proteins are released into the cytoplasm. These capsids associate with the dynein complex which mediates energy dependent retrograde transport to the cell nucleus along microtubules (Topp, K.S., *et al., J. Neurosci. 14*:318-325 (1994); Sodeik, B., *et al.*, "Microtubule and dynein mediated transport of incoming HSV-1 caspids in the nucleus," *21st Intl. Herpes Virus Workshop*, DeKalb, IL

(1996)). Replication-incompetent, recombinant and amplicon HSV-1 vectors expressing the *lacZ* gene have been used to determine the localization and spread of vectors after injection. After single injections into many areas, including caudate nucleus, dentate gyrus and cerebellar cortex, the distribution of β-galactosidase-positive cells was determined (Chiocca, E.A., *et al., N. Biol. 2*:739-746 (1990); Fink, D.J., *et al., Hum. Gene Ther. 3*:11-19 (1992); Huang, Q., *et al., Exp. Neurol. 115*:303-316 (1992); Wood, M., *et al., Exp. Neurol. 130*:127-140 (1994)). Neurons and glia were transduced at the site of injection, and activity was also detected at distant secondary brain areas, in neurons that make afferent connections with the cells in the primary injection site. The retrograde transport to secondary sites is selective to neuroanatomic pathways, suggesting trans-synaptic travel of the virus capsids. Retrograde transport of an amplicon vector has been demonstrated after striatal injections in both the substantia nigra pars compacta and the locus coeruleus (Jin, B.K., *et al., Hum. Gene Ther. 7*:2-15-2024 (1996)). The ability of HSV-1 to travel by retrograde transport to neurons in afferent pathways suggests that the delivery of genes by these vectors can be spread beyond the original injection site to other regions of neuroanatomic importance.

The original report of amplicon-mediated gene delivery to neurons used primary cells in culture (Geller, A.I., and Breakefield, X.O., *Science 241*:1667-1669 (1988)). Amplicon vectors have been used to study neuronal physiology, for example effects of expression of GAP43 or the low affinity nerve growth factor (NGF) receptor on morphology and growth of neuronal cells (Neve, R.L., *et al., Mol. Neurobiol 5*:131-141 (1991); Battleman, D., *et al., J. Neurosci. 13*:941-951 (1993)). Amplicons can direct rapid and stable transgene expression in hippocampal slice cultures (Bahr, B., *et al., Mol. Brain Res. 26*:277-285 (1994)), and this has been used to model both kainate receptor-mediated toxicity (Bergold, P.J., *et al., Proc. Natl. Acad. Sci. USA 90*:6165-6169 (1993)) and glucose transporter-mediated protection of neurons (Ho, D.Y., *et al., J. Neurochem. 65*:842-850 (1995)). *In vivo,* amplicons have been used to deliver

a number of candidate therapeutic genes in different models of CNS diseases. For example, expression of the glucose transporter protects neurons in an induced seizure model (Ho, D.Y., *et al., J. Neurochem. 65*:842-850 (1995); Lawrence, M.S., *et al., Proc. Natl. Acad. Sci. USA 92*:7247-7251 (1995); Lawrence, M.S., *et al., Blood Flow Metab. 16*:181-185 (1996)), *bcl-2* rescues neurons from focal ischemia (Linnik, M.D., *et al., Stroke 26*:1670-1674 (1995)), and expression of TH mediates behavioral changes in parkinsonian rats (During, M.J., *et al., Science 266*:1399-1403 (1994)). Thus, amplicons have proven effective for functional expression of many transgenes in the CNS.

Amplicons have recently been used to generate mouse somatic mosaics, in which the expression of a host gene is activated in a spatial and developmentally regulated fashion. Transgenic mice were engineered with a gennline transmitted NGF gene that contained an inactivating insertional element between the promoter and transcript flanked by the *loxP* sites. The somatic delivery of *cre* recombinase by an amplicon vector successfully activated the expression of NGF in these animals (Brooks, A.I., *et al., Nat. Biotech. 15*:57-62 (1997)). The ability to express genes in specific cells at various points in development will have broad applications, especially for genes for which germline deletion ("knockouts") are conditional lethal mutants.

Traditionally, the stability of transgene expression after transduction, and the cytopathic effect of the helper virus were the limiting features of amplicon-mediated gene delivery into cells of the CNS. Recent advancements have largely addressed these constraints. Several promoter elements, such as preproenkephalin and tyrosine hydroxylase, can drive long-term transgene expression from amplicon vectors when upstream regulatory sequences are included (Kaplitt, M.G., *et al., Proc. Natl. Acad. Sci. USA 91*:8979-8983 (1994); Jin, B.K., *et al., Hum. Gene Ther. 7*:2-15-2024 (1996)). The development of hybrid amplicons containing non-HSV genetic elements that can potentially integrate in a site directed manner (Johnston, K.M., *et al., Hum. Gene Ther. 8*:359-370 (1997)), or form stable replicating episomes (Wang, S. and Vos, J., *J.*

Virol. 70:8422-8430 (1996)), should maintain the-introduced transgene in a genetically stable configuration. Finally, the development of a packaging system devoid of contaminating helper virus (Fraefel, C., *et al., J. Virol. 70*:7190-7197 (1996)) has significantly reduced the cytopathic effects of amplicon vectors in culture and *in vivo*. The easily manipulated plasmid-based amplicon, and the helper virus-free packaging system allows the construction of a virtually synthetic vector which retains the biological advantages of HSV-1, but reduces the risks associated with virus-based gene therapy.

b. Amplicon Vectors for Gene Transfer into Hepatocytes

In another embodiment, the present invention relates to the use of the above-described amplicon vectors for transfer of a torsin nucleic acid molecule into hepatocytes. As discussed in the previous section, HSV-1 amplicon vectors have been extensively evaluated for gene transfer into cells of the nervous system. However, amplicon vectors can also be an efficient means of gene delivery to other tissues, such as the liver.

Certain hereditary liver disorders can be treated by enzyme/protein replacement or by liver transplantation. However, protein infusion can only temporarily restore the deficiency and is not effective for many intracellular proteins. Liver transplantation is limited by donor organ availability and the need for immunosuppression for the lifetime of the patient. Thus, gene transfer to the liver is highly desirable, and consequently, various virus vector systems, including adenovirus vectors (Stratford-Perricaudet, L.D., *et al., Hum. Gene Ther. 1*:241-256 (1990); Jaffe, A.H., *et al., Nat. Genet. 1*:372-378 (1992); Li, Q., *et al., Hum. Gene Ther. 4*:403-409 (1993); Herz, J., and Gerard, R.D., *Proc. Natl. Acad. Sci. USA 90*:2812-2816 (1993)), retrovirus vectors (Hafenrichter, D.G., *et al., Blood 84*:3394-3404 (1994)), baculovirus vectors (Boyce, F.M., and Bucher, N.R.L., *Proc. Natl. Acad. Sci. USA 93*:2348-2352 (1996); Sandig, V., *et al., Hum. Gene Ther. 7*:1937-1945 (1996)) and vectors based on HSV-1 (Miyanohara, A.,

et al., New Biologist 4:238-246 (1992); Lu, B., *et al., Hepatology 21*:752-759 (1995); Fong, Y., *et al., Hepatology 22*:723-729 (1995); Tung, C., *et al., Hum. Gene Ther. 7*:2217-2224 (1996)) have been evaluated for gene transfer into hepatocytes in culture and in experimental animals. Recombinant HSV-1 vectors have been used to express hepatitis B virus surface antigen (HBsAG), *E. coli* P-galactosidase, and canine factor IX (CFIX) in infected mouse liver (Miyanohara, A., *et al., New Biologist 4*:238-246 (1992)). Virus stocks were either injected directly into the liver parenchyma or applied via the portal vein. By either route, gene transfer proved to be highly efficient and resulted in high levels of HBSAG or CFIX in the circulation, and in a large number of P-galactosidase-positive hepatocytes. Although detectable gene expression was transient, a significant number of vector genomes was demonstrated to persist for up to 2 months after gene transfer. The efficiency of longterm gene expression could be increased somewhat by replacing the HCMV *IE1* promoter with the HSV-1 *LAT* promoter to direct the expression of the transgene (Miyanohara, A., *et al., New Biologist 4*:238-246 (1992)).

In two different *ex vivo* studies, primary mouse or human hepatocytes were successfully transduced with HSV-1 amplicon vectors which express *E. coli* β-galactosidase or human growth hormone, respectively (Lu, B., *et al., Hepatology 21*:752-759 (1995); Fong, Y., *et al., Hepatology 22*:723-729 (1995)). After reimplantation of the transduced primary mouse hepatocytes into mouse liver, β-galactosidase-positive cells could be demonstrated for up to two weeks (Lu, B., *et al., Hepatology 21*:752-759 (1995)).

Both recombinant HSV-1 and amplicon vectors can be used for cancer therapy (Breakefield, X.O., *et al., Cancer Gene Therapeutics*, (1995), pp. 41-56). With the goal to enhance the immunogenicity of hepatoma cells and consequently their elimination by host defense systems, Tung, C., *et al., Hum. Gene Ther. 7*:2217-2224 (1996) constructed an amplicon vector that expresses the human interleukin-2 gene (HSV-IL-2). Mouse hepatoma cells (HEPA 1-6) were transduced with either HSV-IL-2 or a control amplicon vector that expressed *E.*

coli β-galactosidase (HSVlac), irradiated, and then used to immunize mice. Animals pretreated in this way were subsequently challenged with intraportal injection of 10^6 viable tumor cells. In the control group (HSVlac), seven out of ten animals developed liver, tumors, whereas in the group of the ten animals pretreated with HSV-IL-2 transduced hepatoma cells, only one animal developed a tumor and that had a much smaller size as compared to those in the control group. Similar amplicon vectors can be used in the present invention.

The present invention is described in further detail in the following non-limiting examples.

Examples

The following protocols and experimental details are referenced in the examples that follow.

Clinical criteria and patient samples

Individuals and families were ascertained from a database of patients diagnosed and treated by members of the Movement Disorders Group at Columbia Presbyterian Medical Center and the Movement Disorders Division at Mount Sinai Medical Center (New York), and through advertisements in the newsletters of the Dystonia Medical Research Foundation. The criteria for the diagnosis of primary torsion dystonia and the method of evaluation were the same as described previously (Bressman, S. B., *et al., Annal Neurol 36*:771-777 (1994)). All subjects gave informed consent prior to their participation in the study.

Two groups of individuals with primary dystonia were considered. The first group consisted of known DYT1 gene carriers from four non-Jewish families previously linked to chromosome 9q34 (Kramer, P., *et al., Am J Hum Gen 55*:468-475 (1994)). These four families were chosen from the seven previously

described families based on individual family lod scores of >+2 at 9q34 markers. Included in this group were also Ashkenazi individuals who carried the founder haplotype of 9q34 alleles (Bressman, S. B., *et al., Annal Neurol 36*:771-777 (1994b)). The second group of individuals had primary dystonia but their linkage status was unknown, i.e. non-Jewish and non-Ashkenazi Jewish individuals from small families and Ashkenazi Jewish individuals who did not have the founder haplotype. This latter group was further subdivided into three clinical subgroups, based on previous studies delineating the DYT1 phenotype (Bressman, S. B., *et al., Annal Neurol 36*:771-777 (1994b); Bressman, S. B., *et al., Neurology 44*:283-287 (1994a); Kramer, P., *et al., Am J Hum Gen 55*:468-475 (1994)). These subgroups are: 1) typical or likely DYT1 phenotype: i.e., early (<28 years) limb-onset with spread of dystonia to at least one other limb, but not to cranial muscles; 2) atypical or unlikely DYT1 phenotype; i.e., focal or segmental cervical-cranial dystonia of any age at onset, or writer's cramp beginning after age 44 years; 3) uncertain DYT1 phenotype; i.e., dystonia not fitting into either of these other categories, such as writer's cramp beginning before age 45, cervical or cranial-onset dystonia spreading down to limbs, or limb-onset spreading up to cranial muscles. Patients with symptoms typical of early onset dystonia were also categorized as uncertain if they had confounding neurological abnormalities.

The four uncertain cases that carried the GAG-deletion are described below. One had a clinical phenotype typical for DYT1, but was classified as uncertain because she had polio as a child, possibly confounding the classification. Another carrier was also typical of DYT1 but was classified as uncertain because of concurrent head tremor and a family history of head and arm tremor. The remaining two carriers had features of typical early onset dystonia; one had early limb onset which spread to other limbs, but also to cranial muscles, and the other had onset in an arm spreading to the neck.

In families of unknown linkage status with multiple affected members a single classification was assigned to all affecteds within the family. If families had members with both uncertain and typical phenotypes, a classification of

typical was assigned; if families had members with both uncertain and atypical phenotypes, a classification of atypical was assigned; and if families had members with both atypical and typical phenotypes, a classification of uncertain was assigned.

Individuals with a wide range of ethnic and geographic ancestry were also sought to be included. This study was approved by the review boards of both institutions.

DNA isolation, lymphoblast lines and Southern blots

Venous blood samples were obtained from participating individuals. DNA was extracted from whole blood (Gusella, J., *et al., Proc Natl Acad Sci USA 76:*5239-5242 (1979)) or from lymphoblast lines established from blood lymphocytes by EBV transformation (Anderson, M., & Gusella, J. *In Vitro 29*:856-858 (1984)). CEPH pedigree DNA was obtained from CEPH (Centre d'Etude du Polymorphisms Humain, Paris, France). For Southern blots genomic DNA was digested with PstI, HindIII and EcoRI (NEB) according to the manufacturer's instructions. Digested DNA was resolved on 1% agarose gels at 70V for 16 h. Southern blotting was performed using standard techniques and the filters were hybridized to cDNAs (see below).

Isolation of RNA, northern blots, RT-PCR

Cytoplasmic RNA was isolated from lymphoblasts and fibroblasts established from patients and controls (Sambrook, J., *et al.*, in *Molecular Cloning. A Laboratory Manual,* Cold Spring Harbor Press, Cold Spring Harbor, NY (1989)). Total RNA was extracted from human adult and fetal tissue obtained at autopsy using the guanidinium thiocyante method (Chirgwin, J. M., *et al., Biochem 18*:5294-5300 (1979)). Tissues were obtained from both control and DYT1 carrier individuals and included brain cortex, cerebellum, hippocampus, lung, liver, muscle, placenta, spleen, thyroid, intestines, and eye. A northern blot was prepared from this RNA following standard procedures

(Sambrook, J., *et al.*, in *Molecular Cloning. A Laboratory Manual,* Cold Spring Harbor Press, Cold Spring Harbor, NY (1989)). In addition, northern blot filters containing 2 μg of poly (A+) RNA from 8 different adult human tissues and four different fetal human tissues were used (Clontech). First strand cDNA synthesis was performed on lymphoblast RNA samples with oligo dT and random primers using Superscript II reverse transcriptase (Gibco;Newman, P. J., *et al.*, *J Clin Invest 22:*739-743 (1988)). The reactions were carried out at 42°C for 90 min followed by gene specific PCR amplification to generate the various cDNAs in the critical DYT1 region from patients and controls.

Cosmid contig

Cosmids were isolated from two human chromosome 9-specific libraries: the Lawrence Livermore library, which was constructed in Lawrist 16 using DNA from a chromosome 9-only somatic cell hybrid (Van Dilla, M. A., & Deaven, L. L. *Cytometry 11:*208-218 (1990)); and the Los Alamos library, which was constructed in sCos (Stratagene) from flow-sorted metaphase human chromosomes (Deaven, L. L., *et al.*, *Symp Quant Biol 51:*159-167 (1986)). Cosmid colony grid filters were stamped and prepared for hybridization as described (McCormick, M. K., *et al.*, *Genomics 18:*553-558 (1993); Murrell, J., *et al.*, *Genomics 25:*59-65 (1995)). Filters were screened with gel-purified YAC DNA from a 400 kb critical region and over 800 positive colonies were picked, gridded and stamped for hybridization. A cosmid walk was initiated from both ends of the critical region starting with the end clone of cosmid LL09NC0150H11 and several D9S63 positive cosmids (18D5LA, 37H5LA). End sequences of hybridizing cosmids were used to generate PCR primers to continue the walk by re-screening grids. The resulting set of about 60 cosmids was digested with EcoRI and fragments were resolved by agarose gel electrophoresis to distinguish similar and novel regions. An overlapping subset of 11 cosmids was then aligned by digestion with EcoRI, XhoI and NotI in a series of single and double digests. Fragments were resolved by electrophoresis in 1% agarose gels, transferred to

Southern blots and hybridized to cosmid ends, exons and unique sequence in the region, generated by PCR or with synthetic 20 bp oligonucleotides.

Hybridization

Probes (gel-purified YAC DNA, cosmid ends, exon clones, PCR products and cDNAs) were labeled by random priming (Feinberg, A. P., & Vogelstein, B., *Anal Biochem 137*:266-267 (1984)) using [^{32}P]dATP (3000 Ci/mmol; NEN). Oligonucleotides were end-labeled with T4 polynucleotide kinase (NEB) using [^{32}P]dATP (6000 Ci/mmol; NEN). Probes were preannealed with CotI DNA, human placenta DNA and vector DNA, as necessary to saturate repeat sequences.

Hybridizations were performed in Church-Gilbert buffer at 55°C overnight. Northern filters were hybridized in 5X SSPE, 50% formamide, 5X Denhard's solution, 0.5% SDS and 300 μg/ml salmon sperm DNA overnight at 42°C. Filters were washed and exposed to autoradiographic film as described (Murrell, J., *et al., Genomics 25*:59-65 (1995)).

Exon amplification, cDNA library screening and extension

Exon amplification was performed on cosmids spanning the region using vectors pSPL1 (Buckler, A. J., *et al., Proc Natl Acad Sci USA 88*:4005-4009 (1991)) and pSPL3-IV (Church, D. M., *et al., Nat Genet 6*:98-105 (1994)). RT-PCR products were digested with BstX1 to eliminate vector-only products (Church, D. M., *et al., Nat Genet 6*:98-105 (1994)). Cloned exon fragments were used to screen human fetal and adult cDNA libraries by colony hybridization (Sambrook, J., *et al.*, in *Molecular Cloning. A Laboratory Manual,* Cold Spring Harbor Press, Cold Spring Harbor, NY (1989)). Libraries were prepared in ZAP by Stratagene and included adult human striatum, hippocampus, substantia nigra, caudate putamen, brainstem, heart, spleen and liver, and fetal human brain and retina. The sequences generated from these cDNA clones were aligned and edited using the Sequencher program (Gene Codes). cDNA contigs were extended in two ways - by rescreening libraries with PCR fragments generated from the ends

of the contig; and by using 5' and 3' RACE and MARATHON PCR systems (Frohman, M. A., *et al., Proc Natl Acad Sci USA 85:*8998-9002 (1988); Apte, A. N., & Siebert, P. D., *Biotechniques 15:*890-893 (1993)) as modified by Clontech.

Sequencing

Dideoxysequencing was performed using the Sequitherm Long Read Cycle Sequencing Kit (Epicenter Technologies) either with infrared labeled vector primers for the LICOR sequencing machine or with specific primers labeled with ^{33}P-dATP (2000 Ci/mmol, NEN) for standard sequencing. Direct sequencing of PCR products was done using an enzymatic cleanup process with exonuclease I and shrimp alkaline phosphatase (USB) for 15 min at 37°C and 15 min at 85°C, followed by sequencing with Sequenase (USB). The LICOR sequence was read using the BaseImagIR software package (LICOR) which includes data collection and image analysis software. The ^{33}P-dATP gels were transferred to 3 MM Whatman, dried and exposed to autoradiographic film overnight and then read and entered manually into the GCG programs (Wisconsin Package Version 9.0, Genetics Computer Group, Madison, WI).

SSCP analysis

DNA sequences were screened for mutations by PCR of 100 - 300 bp fragments followed by SSCP analysis, using first strand cDNAs synthesized from patient and control lymphoblast RNA and genomic DNA. PCR reactions were performed as described in Ozelius, L., *et al., Am. J. Hum. Genet. 50:*619-628 (1992) in a total reaction volume of 10 μl. SSCP analysis of the PCR amplification products was carried out as described (Orita, M., *et al.*, *Genomics 5:*874-879 (1989); Hayashi, K., & Yandell, D. W., *Hum Mutation 2:*338-346 (1993)). All fragments with altered migration were sequenced directly and evaluated in families to check inheritance, and in controls to determine whether they represented normal polymorphisms. When single base pair changes altered restriction sites, restriction digestion of PCR products was used to replace SSCP

analysis. In these cases, standard PCR was performed in a 25 μl final volume without radioactivity. The PCR products were digested according to manufacturer's instructions (NEB) and visualized by staining with ethidium bromide after electrophoresis in 2-3.5% agarose gels.

Example 1
Cosmid contig and transcript map

A cosmid contig was constructed across the 150 kb target region on chromosome 9q34 between polymorphic markers, D9S2161 and D9S63 to facilitate identification of genes. Gridded arrays of cosmids from two chromosome 9-specific libraries were screened initially with four YACs: 8H12, 183D9, 251H9 and 22A4, spanning this region. A positive subset of cosmids was screened sequentially with end sequences from cosmids, starting on the centromeric side with the end of cosmid LL09NC0150H11, which hybridized to YAC 8H12, and, at the telomeric end, with D9S63-positive cosmids, 37F5LA and 18D5LA, which hybridized to the other three YACs. New overlapping cosmids were compared by restriction digestion and gel electrophoresis, and end sequences were obtained from novel fragments and used to rescreen secondary cosmid grids. A cosmid contig with over 3-fold redundancy was generated across the genomic region between LL09NC0150H11 and 37F5LA. A restriction map was constructed using a representative subset of 11 overlapping cosmids by determining fragment sizes following gel electrophoresis of single and double digests using the restriction enzymes, EcoRI, XhoI and NotI. Fragments were aligned by size and hybridization patterns using markers D9S2161 and D9S63, cosmid end clones, cloned exons, and oligonucleotides from unique regions (Fig. 1). The estimated length of the contig between the defining markers is 150 kb.

Genes in this region were identified by exon amplification, which allows cloning of exons by virtue of flanking splice sites in genomic DNA during cellular processing of RNA expressed by splicing vectors (Buckler, A. J., *et al.*, *Proc Natl Acad Sci USA 88*:4005-4009 (1991); Church, D. M., *et al.*, *Nat Genet*

6:98-105 (1994)). A subset of cosmids from the critical region were digested with PstI or SacI, or with both BamHI and BglII, and cloned into these vectors. Of over 60 putative exons trapped, 28 produced independent sequences. These exon clones were then used to screen human cDNA libraries from different human adult and fetal tissues, which had been generated by priming with oligo dT or random primers. Five cDNAs were represented in multiple overlapping clones: DQ1 from fetal brain, adult frontal cortex and adult liver; DQ2 from adult substantia nigra, hippocampus and frontal cortex; DQ3 from adult frontal cortex; DQ4 from adult frontal cortex and fetal brain; and DQ5 from adult occipital cortex, substantia nigra and frontal cortex. All but three of the 28 unique putative exons were accounted for by these cDNAs. Hybridization of these three to northern blots from a number of adult human tissues (Clontech) revealed no corresponding message species. Further screening of cDNA libraries with these putative exons did not yield positive clones. Therefore these three may be the result of cryptic splice junctions or may represent low abundance messages.

The five cDNAs were extended in both directions by 5' and 3' RACE (Frohman, M. A., *et al., Proc Natl Acad Sci USA 85*:8998-9002 (1988)) and sequenced in multiple clones. Transcripts were then aligned across the cosmid contig by hybridizing Southern blots of restriction digested cosmid DNA with 5' and 3' cDNA ends, exon clones and oligonucleotides corresponding to cDNA sequence (Fig. 1). Estimated genomic regions covered by these genes are: 8 kb for DQ1; 13 kb for DQ2; 10 kb for DQ3; 52 kb for DQ4 and >40 kb for DQ5. Since only the 3' untranslated region of cDNA DQ5 overlapped the critical interval, this gene was excluded from this study. Given the extensive exon trapping carried out in this region, and the fact that 2-19 exons were identified for each cDNA, it is possible that these transcripts account for all the genes in the critical region. However, several large regions (> 10 kb each), one centromeric to DQ1 and two within the first two introns of DQ4, could contain other genes, particularly genes with one or no introns that would be missed by exon amplification.

Example 2
Sequence of cDNAs

DQ1 (encoding a torsinB sequence, SEQ ID NO:4) and DQ2 (encoding a torsinA sequence, SEQ ID NO:2) are highly homologous transcripts with 72% identity at the nucleotide level and 69% amino acid identity in the predicted protein sequence. The genes are in opposite orientations with their 3' ends <12 kb apart in the genome (Fig. 1). The longest transcript for DQ2 consists of 2,072 bp with a predicated open reading frame of 998 bp from nucleotide 43 to 1041 of Fig. 2 (SEQ ID NO:5). A composite nucleotide sequence of the torsin A gene comprised of genomic and cDNA sequence is set forth in SEQ ID NO:1. A composite nucleotide sequence of the torsin A gene comprised of cDNA sequence is set forth in SEQ ID NO:5.

The sequence around the putative ATG translation start site contains the critical -3 purine residue, but none of the other features of the Kozak, M., *Nucleic Acids Res 15:*8125-8148 (1987) consensus sequence. There is an in-frame termination codon in overlapping genomic sequence from cosmid 23G1LA starting 365 bp upstream of the 5' end of this cDNA. The 3' untranslated portion is 1031 bp long and contains two poly A^+ addition sites, one, ATTTAAA, at nucleotide 1390 and the other, AATAAA, at nucleotide 2054, with poly A tails present about 20 bp down from each of them in several of the cDNA clones.

The cloned DQ2 cDNA appears to be essentially full length based on the sizes of corresponding transcripts. Clone H4, ATCC Accession No. 98454 contains nucleotides 99 to 1377 as set forth in SEQ ID NO:5. Northern blots of adult and fetal human RNA showed two ubiquitous messages of about 1.8 kb and 2.2 kb which hybridized to a probe from the coding portion of the cDNA (Fig. 3A). Only the 2.2 kb message hybridized to sequences 3' to the first poly A^+ addition site, indicating that the larger species may represent utilization of the second poly A^+ addition site. In fetal brain, lung and kidney, as well as adult brain, heart and pancreas, another message species of about 5 kb was present in low abundance. The 1.8 kb and 2.2 kb messages, and no novel species, were seen

in autopsy tissues from AJ individuals bearing the founder mutation, including adult lymphoblasts, fibroblasts and cerebellum, as well as fetal brain, muscle, spleen, intestines, eye, lung and liver. The open reading frame of transcript DQ2 predicts a 6.81 pI polypeptide of 332 amino acids with a calculated molecular weight of 37,813 D, which is termed torsinA (Fig. 4)(SEQ ID NO:2).

cDNA DQ1 is 2504 bp with an open reading frame of 802 bp (Fig. 2) (SEQ ID NO:6 and 7) and was deposited at the ATCC as accession no. 98455. Based on the strong similarities between cDNAs DQ1 and DQ2, and between genomic sequence 5' to cDNA DQ1 in cosmid 54A5LA and the 5' coding end of cDNA DQ2, the methionine start site for the DQ1 message is probably not in the existing DQ1 clone. The nucleotide sequence set forth in SEQ ID NO:3 and the corresponding amino acid sequence set forth in SEQ ID NO:4 were generated from a combination of genomic and cDNA sequence. The 3' untranslated portion is 1702 bp long with a poly A^+ addition site, AATAAA, at position 2483 and a poly A tail about 20 bps downstream (SEQ ID NO:6). Northern blot analysis revealed a ubiquitously expressed message of about 2.8 kb, present at low levels in adult brain, but not detectable in fetal brain (Fig. 3B). The open reading frame of the existing clone encodes 290 amino acids, suggesting that the corresponding protein, which we have named torsinB, has a molecular weight > 32,000 D (Fig. 4) (SEQ ID NO:4).

Example 3
Mutational analysis

Two possible mechanisms of mutation were considered in this dominant disorder: disruptive mutations, which would inactivate the protein and result in haploinsufficiency of the encoded protein; and missense mutations, which would cause a "gain-of-function" or "dominant negative" effect of the mutant form of the protein that would override the function of the normal protein or interfere with other proteins.

Gross alterations in sequence were excluded by Southern blot analysis using genomic DNA from 30 dystonia patients from different ethnic backgrounds hybridized to the four cDNAs. Transcripts from the critical region were then screened for sequence variations using lymphoblast RNA from 14 individuals affected with torsion dystonia from different families representing 12 unique haplotypes in an extended region surrounding the DYT1 gene (D9S62a to ASS) and two control individuals, one AJ and one non-Jewish. Of these 12 unique haplotypes, four were from families that demonstrated clear chromosome 9 linkage. Theses included four non-Jewish families (Kramer, P., *et al.*, *Am J Hum Gen 55:*468-475 (1994)), two of which were French-Canadian and shared a common haplotype and the AJ founder mutation. The four distinct disease-linked haplotypes in these six families were initially assumed to represent independent mutations in DYT1. After initial RT-PCR with primers in the 3' and 5' end of the transcripts, nested PCR was carried out in overlapping fragments of 150-300 bp. Fragments were resolved by SSCP analysis and all variant bands were sequenced in both directions. All transcripts showed a number of variations in both coding and noncoding regions (Fig. 2). Most of these were "silent", involving single base pair substitutions in the 5' and 3' untranslated regions or in the third position of triplet codons, such that the encoded amino acid would not be altered. Only three of these changes affected the amino acid sequence: 1) valine isoleucine in DQ4; 2) aspartic acid histidine in DQ2; and 3) deletion of a glutamic acid in DQ2. All sequence variations in the coding regions were analyzed in genomic DNA after determining the exon/intron structure in cosmid DNA. All of these nucleotide changes, except one - the GAG-deletion in DQ2 (SEQ ID NO:5 at nucleotide positions 946-948), were confirmed as polymorphisms by their presence in control samples. Surprisingly, the GAG-deletion in DQ2 was present in all six individuals representing the four confirmed DYT1-linked haplotypes.

To pursue this finding, the co-segregation of the GAG-deletion with carrier status in all known chromosome 9-linked families was examined, as well

as in a large number of AJ and non-Jewish controls. This GAG-deletion was analyzed using PCR products generated from genomic DNA samples in three ways: SSCP, direct sequencing, or digestion with BseRI, which cuts 10 bp downstream of the normal GAGGAG sequence, but does not cut the GAG-deletion sequence (Fig. 5). The association of the GAG-deletion with carrier status in the chromosome 9-linked families was complete. All 261 affected and unaffected obligate gene carriers in 68 chromosome 9-linked families were heterozygous for this deletion (including 64 AJ families carrying the founder haplotype and 4 non-Jewish families) (Table 1). Strikingly, the deletion was not present in 260 AJ and 274 non-Jewish control chromosomes, and was never observed in the homozygous state in any individual.

To further assess the role of this deletion in primary dystonia, an additional 155 individuals with varying clinical manifestations from families which were too small for linkage analysis were typed. The association with the typical manifestation of early onset torsion dystonia was very strong in these families. Among 41 cases of typical early onset dystonia from 19 different families, all affected individuals in14 of the families carried the GAG-deletion, while affected individuals from five other families did not. Affecteds from two of these five families are suspected to have dopa-responsive dystonia, which closely mimics the phenotype of DYT1, but were unconfirmed. Some typical cases that do not carry the GAG-deletion may have other, as yet unidentified, mutations in DQ2, or mutations in other genes; for example, features of the DYT1 phenotype occur in some individuals who carry a mutation on chromosome 8. Among 38 cases (from 11 AJ, 27 non-Jewish families), for which the diagnosis of early onset dystonia was uncertain, four carried the GAG-deletion (1 AJ, 3 non-Jewish) while the remaining 34 did not. (For clinical description of these four uncertain carriers see-Methods.) Among the 76 individuals classified as atypical, that is not having features typical of early onset dystonia (36 AJ, 2 Sephardic Jewish [SJ] and 38 non-Jewish families), none carried the GAG-deletion. Collectively, these observations provide compelling

evidence that the GAG-deletion is responsible for the vast majority of cases of typical early onset dystonia. These include individuals of AJ descent, in whom the founder mutation causes >95% of cases (Risch, N., *et al.*, *Nature Genetics* 9:152-159 (1995)), as well as most non-Jewish cases of varied ethnic backgrounds.

The identification of a single mutation (GAG-deletion) on affected chromosomes responsible for almost all cases of typical early onset dystonia is remarkable. Two possible explanations may account for this surprising finding: 1) all these cases may be ancestrally related, representing a unique founder mutation which predates the introduction of this mutation into the AJ population; or 2) the same mutation has arisen independently, and is the only change (or perhaps one of a few) that can result in the early onset dystonia phenotype. To distinguish between these possibilities, three polymorphic sequence variations (A, B, and C, Table 2) were identified in a 5 kb region surrounding the GAG-deletion and used these to perform a more detailed allelic analysis. In affected individuals carrying the four confirmed DYT1-linked haplotypes, two different patterns of alleles were observed, 1, 1, 2 on three of the disease-bearing chromosomes (families 5, 9, and 16) and 1, 2, 1 on one of the disease-bearing chromosome (family 1, Table2). This finding clearly supports the idea that the same mutation (GAG-deletion) has arisen more than once. Sixty AJ and 60 non-Jewish control individuals were also typed with these markers and the frequencies determined for the different allele patterns. Among the controls, only three patterns were observed at loci A, B, C: 1, 1, 2 (AJ controls = 68%; non-Jewish controls = 55%); 1, 2, 1 (AJ = 20%; non-Jewish = 34%); and 2, 1, 2 (AJ = 12%; non-Jewish = 11%), suggesting that these polymorphisms are in strong linkage disequilibrium with each other. The high frequency of the 1, 1, 2 pattern in GAG-deleted chromosomes in 9-linked families is consistent with the high incidence of this pattern in controls and is not in conflict with independent mutations. When the other 18 GAG-deleted chromosomes in patients with unknown linkage status were typed for A, B and C alleles- 15 carried 1, 1, 2 (families 2-4, 6-8, 10-15,

and 17-22), one had the 1, 2, 1 pattern (family 2) and two others (families 3, 4) were unphaseable but likely have the 1, 2, 1 pattern (see below) (Table 2). These observations support the conclusion of at least two independent mutations but the high frequency of the 1,1,2 pattern in the normal population might mask our ability to distinguish additional independent events.

To examine whether the individuals bearing identical allelic patterns at markers A,B, and C showed additional evidence of common ancestry, their haplotypes were investigated at flanking microsatellite markers (D9S159, D9S2160, D9S2161, D9S63 and D9S2162). This analysis revealed that the four individuals carrying the uncommon (1, 2, 1) pattern at A, B, C are likely to share a common ancestry (families 1-4, Table 2). All of these families come from the same region of North Carolina and Virginia, have British ancestry, and affected individuals share alleles across ≥200 kb surrounding the GAG-deletion. Examination of the remaining 18 chromosomes, all of which carry the same (1, 1, 2) haplotype, also reveals some interesting commonalities that fall into two categories. First, three families share a portion of the alleles characteristic of the AJ founder mutation. The SJ individual (family 6, Table 2) carries a haplotype at four loci which is identical to the Ashkenazi founder haplotype (family 5). A natural question is whether the disease chromosome in this Sephardic individual was introduced recently through an unidentified Ashkenazi ancestor, or alternatively whether this progenitor mutation existed in the Jewish population prior to the separation of the Sephardim from the Ashkenazim approximately 1,000 years ago. The affected individual in family 7, reported to be non-Jewish of mixed English and Austria-Hungarian ancestry, was consistent with carrying the founder AJ chromosome across the whole 500 kb haplotype region. It is likely that this individual has inherited the DYT1 gene from a recent Ashkenazi ancestor. A third, possibly related chromosome was found in an Italian family (family 8). This individual shared the AJ founder alleles ≥90 kb centromeric to the GAG-deletion. This raises the intriguing possibility of a Mediterranean origin for this mutation predating its introduction into the Ashkenazi Jewish population.

Alternatively, as before, the chromosome could be of more recent Ashkenazi ancestry. Second, three common haplotypes are found in other families, distinct from the AJ founder haplotype: 1) family 9 (German origin) shared the haplotype across ≥420 kb with family 10 (Irish origin), and apparently with family 11 (Ashkenazi Jewish; although was unphased). The long stretch of DNA shared by these individuals suggests a common origin, despite the varied ethnicity. 2) Two of the French-Canadian families (12 and 13) also bear a shared haplotype over a smaller region, ≥130 kb. Thus, there may be shared ancestry here as well, but in the more distant past. 3) Another two families, 14 (German origin) and 15 (Irish origin) shared the same haplotype across ≥320 kb. Among the remaining families, there appears to be an additional seven haplotypes. In total, there appears to be at least 12 distinct classes of haplotypes, suggesting that the same mutation has occurred at least that many times, however, it remains possible that all of the chromosomes with the 1,1,2 pattern are ancestrally related in the very distant past.

TABLE 4: Genotype of GAG deletion in candidate cDNA in affected individuals and controls

Categories		Families	Genotype in Individuals		
			+/+[a]	+/-	-/-
Controls:					
AJ		130	130	0	0
NJ		137	137	0	0
Affected and obligate carriers[b] in 9-linked families:					
AJ founder haplotype		64	0	173	0
NJ		4	0	88	0
Affecteds of unknown linkage[c]:					
AJ:	typical[d]	1	0	4	0
	uncertain	11	10	1	0
	atypical	36	36	0	0
SJ:	typical	1	0	2	W0
	atypical	2	2	0	0
NJ:	typical	17	5	30[e]	0
	uncertain	27	24	3	0
	atypical	38	38	0	0

[a] +: GAGGAG, -: GAG (del)

[b] lod score > +2 for 9q34 markers (Kramer *et al.*, *Am. J. Hum. Gen. 55*:468-475 (1994))

[c] families too small for linkage analysis

[d] see clinical criteria in methods for definitions

[e] from 12 families

TABLE 5: Haplotypes surrounding GAG-deletion

				<40kb	70kb	100kb	<40kb	50kb	5kb		100kb	130kb	
	Ethnicity[a]	Family	Status[b]	D9S62a	D9S62b	D9S159	D9S2160	D9S2161	A[c]	B *	C	D9S63	D9S2162
	Br	1	T	4	4	12	5	5	1	2	1	2	5
	Br	2	T		4	12	3/5	5	1	2	1	2	5
5	Br	3	T				5	5	1	1/2	1/2	2	
	Br	4	T			16/18	5/6	3/5	1/2	1/2	1/2	0/2	¼
	AJ	5	T	2	8	12	4	5	1	1	2	16	4
	SJ	6	T	4	10	6	4	5	1	1	2	16	4
	A-H/Br	7	U	2/6		4/12	4	2/5	1	1	2	-2/16	2/4
10	It	8	T	2	8	14	4	5	1	1	2	0	
	G	9	T	4	6	16	5	4	1	1	2	14	5
	I	10	T		6	16	5	4	1	1	2	14	4
	AJ	11	U	0/4	6/10	14/16	5	4/5	1	1	2	16/18	2/5
	FC	12	T	4	6		6	4	1	1	2	0	2
15	FC	13	U				4/6	3/4	1	1	2	0/14	

TABLE 5: Haplotypes surrounding GAG-deletion

			<40kb		70kb	100kb	<40kb	50kb	5kb		100kb	130kb
Ethnicity[a]	Family	Status[b]	D9S62a	D9S62b	D9S159	D9S2160	D9S2161	A[c]	B *	C	D9S63	D9S2162
G	14	T	2	0	16	5	2	1	1	2	6	5
I	15	T			4	5	2	1	1	2	6	5
FC	16	T	4	8	14	4	2	1	1	2	14	4
A	17	T				5/6	5/6	1	1	2	10/16	
Af-A	18	U				4	1/4	1	1	2	12/16	
Sw	19	T			12/16	3/4	4		1/2	1/2	2/6	2/3
I	20	T	4	-10/12		5	3/5	1	1	2	0/16	
Br	21	T	4	10		5	1/5	1	1	2	20	
AJ	22	T	4	6	12	3	5	1	1	2	18	5

Bold = Chromosome 9 linked families
shaded = shared haplotype regions

[a] AJ = Ashkenazi Jewish; A-H = Austrian - Hungry; AF - Am = African American; SJ = Sephardic Jewish; I = Irish; FC = French Canadian; A = Arab; Sw = Swedish; G = German; Br = British; It = Italian;

[b] T = typical; U = uncertain; see clinical criteria in methos for definitions.

[c] Markers A, B, and C are single base pair polymorphisms surrounding the GAG-deletion (indicated by*). A is detected by loss or retention of an Nla III site; B is an SSCP shift; and C is detected by loss or retention of a TaqI site. The individual frequencies of these alleles were determined in 60 AJ and 60 NJ chromosomes: A = AJ - 90%, 1 and 10%, 2; NJ - 88%, I and 12%, 2; B = AJ - 81%, I and 19%, 2; NJ - 66%, 1 and 34%, 2: C = AJ - 19%, 1 and 81%, 2; NJ - 35%, 65%, 2.

Discussion

Positional cloning was used to identify a strong candidate for the DYT1 gene on human chromosome 9q34 which is responsible for the early onset form of primary dystonia. Mutational analysis revealed a 3-bp deletion in the coding portion of a transcript which was the only non-polymorphic change identified on disease-bearing chromosomes. This mutation was uniquely associated with typical cases of early onset dystonia and appears to have arisen independently on different haplotypes in a number of ethnic groups. Thus, apparently only one, or one of a few variations in the encoded protein can give rise to this particular phenotype. The deletion results in loss of one of a pair of glutamic acid residues near the carboxy terminus of a novel protein, termed torsinA. These glutamic acids and flanking cysteine residues are conserved in an adjacent, homologous human gene, encoding a protein termed torsinB, and in related mouse and rat sequences. TorsinA and B define a new family of ATP-binding proteins with a distant relationship to the HSP100/Clp family of proteins (Schirmer, E. C., *et al.*, *TIBS 21*:289-296 (1996)).

Insights into the protein relationships and possible function of torsinA were revealed by searches of GenBank protein databases. A closely related deduced protein sequence is encoded by cosmids from *Caenorhabditis elegans*, here termed, torsin-related protein in C. elegans (torpCel). ESTs were also identified corresponding to human, mouse and rat torsinA and torsinB, as well as to two other related proteins, torp1 and torp2. Figure 4A shows an amino acid alignment of these predicted proteins. The glutamic acid pair, bearing the deletion in affected individuals, is conserved in all human, rat and mouse torsinA and torsinB transcripts, suggesting it is part of a functional domain. Although the glutamic acid pair is absent in the torps, the neighboring residues are fairly well conserved, including the cysteine residues which flank this region. A phylogeny analysis suggests that torsinA and torsinB are most closely related to each other (~70% amino acid sequence identity), and that they and the torps are equally distant (~50% identity).

A more distant similarity to known proteins offers some insight into potential function: the middle ~200 amino acids of the torsins and torps are similar to a conserved domain in the HSP100/Clp family of proteins (Schirmer, E. C., *et al.*, *TIBS 21*:289-296 (1996); Perier, F., *et al.*, *Gene 152*:157-163 (1995)). Members of the HSP100/Clp family have chaperone functions or proteolytic activity, which can confer thermotolerance, allow correct folding of proteins and regulate protein signaling (Parsell, D. A., & Lindquist, S., *Annu Rev Genet 27*:437-496 (1993)). HSP100/Clp proteins are distinguished by two features: they bind ATP and/or have ATPase activity; and they occur in oligomeric complexes with one or more additional protein species. Figure 4C compares the torsin family to two representative members of the HSP100 protein family: HSP101, a heat shock protein from soybeans in the HSP100 Subfamily 1B; and SKD3, a ubiquitous mouse protein in the Subfamily 2M (Perier, F., *et al.*, *Gene 152*:157-163 (1995); Schirmer, E. C., *et al.*, *TIBS 21*:289-296 (1996)). The most robust feature is a conserved ATP/GTP-binding sequence comprising two motifs: the nucleotide-binding site "A" (GxTGxGKT/S) followed ~60 amino acids later by the Mg^{++}-binding site "B" (ShhhFDEhEKxH), where x indicates variable residues and h indicates hydrophobic residues (Walker, J. E., *et al.*, *EMBO J 1*:945-950 (1982); Confalonieri, F., & Duguet, M., *Bioessays 17*:639-650 (1995)). In a conserved stretch of 140 amino acids that include the nucleotide binding domain, torsin family members are 25-30% identical to HSP100 family proteins. Key residues of the HSP100 consensus site IV (Schirmer, E. C., *et al.*, *TIBS 21*:289-296 (1996)) and another site (SN; Fig. 4C) are also conserved, but consensus site V is absent in the torsins. Interestingly, a mutation in the carboxy region of the *E. coli* HSIV/ClpY in this family blocks binding to its companion protein (Missiakas, D., *et al.*, *EMBO J 15*:6899-6909 (1996)). The discrete glutamic acid deletion in the carboxy end of torsinA would be consistent with the ability of the mutant protein to interfere with binding and activity of other subunits. It is tempting to consider torsinA in the same

superfamily as the heat shock/proteolytic proteins, as it would be consistent with a dominant-negative effect mediated by disruption of a multimeric complex.

There are few other clues from the deduced protein sequences of this apparently new class of proteins that yield insight into their function. The forty-one amino acids at the putative N-terminal of torsinA consist of two 20 amino acid hydrophobic domains. The first of these begins with two basic amino acids, and ends with a polar and an acidic amino acid; it fulfills the criteria of a leader sequence for a transmembrane or membrane translocated protein (Boyd, D., & Beckwith, J., *Cell 62*:1031-1033 (1990)). There are several possible phosphorylation sites which are conserved in both torsinA and torsinB: two for protein kinase C, and one for casein kinase II; as well as a number of putative N-myristylation sites (Prosite analysis). Six cysteine residues are conserved with other torsin family members (Fig. 4B).

The finding of the same 3-bp mutation in the heterozygous state in most cases of typical early onset dystonia is surprising, yet robust. There are only a few examples of recurrent mutations which cause dominantly inherited conditions. These include: loss of a positively charged arginine in the fourth transmembrane helix of the $_1$ subunit of the L-type voltage sensitive calcium channel, which is the only type of mutation found to cause hypokalemic periodic paralysis (Grosson, C. L., *et al., Neuro. Disord 6*:27-31 (1995); Fontaine, E., *et al., Nat Genet 6*:267-272 (1994)); a glycine to arginine substitution in the transmembrane domain of the fibroblast growth factor receptor-3 (FGFR3) seen in almost all cases of achondroplasia (Bellus, G. A., *et al., Am J Hum Genet 56*:368-373 (1995)); common missense mutations seen in hypertrophic cardiomyopathy (Watkins, H., *et al., Am J Hum Genet 53*:1180-1185 (1993)); and CAG expansions in the coding regions of a number of genes causing neurodegenerative diseases (for review see Gusella, J., *et al., Proc Natl Acad Sci USA 76*:5239-5242 (1979); Paulson, H. L., & Fishbeck, K. H., *Ann Rev Neurosci 19*:79-107 (1996)). In all these cases it appears that the same mutations occur repeatedly as independent events, while other mutations in the same gene

apparently cause a different syndrome, have no phenotype, or are incompatible with life.

Early onset dystonia (DYT1) represents the most severe and most common form of hereditary dystonia. This and other genetic forms of dystonia usually follow an autosomal dominant pattern of inheritance with reduced penetrance (30-40%). Six genes causing non-degenerative forms of dystonia have been mapped on human chromosomes: dopa-responsive - dominant on 14q21-22 (Nygaard, T., *et al.*, *Nature Genetics 5:*386-391 (1993); Endo, K., *et al.*, in *Monographs in Neural Sciences Age-related dopamine-dependent disorders*, Segawa, M & Nomura, Y., eds., Karger Publishers, New York (1995), pp. 120-125) and recessive on 11p11.5 (Ludecke, B., *et al.*, *Hum Genet 95:*123-125 (1995); Knappskog, P. M., *et al.*, *Hum Mol Genet 4:*1209-1212 (1995)); paroxysmal on 2q (Fink, J. K., *et al.*, *Am J Hum Genet 59:*140-145 (1996); Fouad, G. T., *et al.*, *Am J Hum Genet 59.*135-139 (1996)); a late onset, focal on 18p (Leube, B., *et al.*, *Hum Mol Genetics 5:*1673-1678 (1996)); a mixed phenotype on 8; and the DYT1 gene on chromosome 9q34 (Ozelius, L., *et al.*, *Neuron 2:*1427-1434 (1989)). Two of these other dystonia genes have been identified and both implicate decreased dopaminergic transmission in dystonia. Dopa-responsive dystonia can be caused by disruption of tyrosine hydroxylase, the rate limiting enzyme in dopa synthesis (Ludecke, B., *et al.*, *Hum Genet 95:*123-125 (1995); Knappskog, P. M., *et al.*, *Hum Mol Genet 4:*1209-1212 (1995)) and by haploinsufficiency of GTP cyclohydrolase I, needed for synthesis of the tyrosine hydroxylase cofactor, biopterin (Furukawa, Y., *et al.*, *Adv Neurol 69:*327-337 (1996)). The only genetic rodent models reported to date - the dystonic (dt) rat (LeDoux, M. S., *et al.*, *Brain Res 697:*91-103 (1995)), the mouse mutant, dystonia musculorum (dMd) (Brown, A., *et al.*, *Nat Genet 10:*301-306 (1995)); and hamsters with paroxysmal dystonia (dt52) (Nobrega, J. N., *et al.*, *Neurosci 64:*229-239 (1995) and Nobrega, J. N., *et al.*, *Neurosci 71:*927-937 (1996); Pratt, G. D., *et al.*, *J Neurochem 1995:*2153-2158 (1995)) - do not match the genetic or neurobiologic features of human dystonias.

Although there is no distinctive neuropathology in primary dystonia (Hedreen, J. C., *et al., Adv Neurol 50:*123-132 (1988); Zeman, W., & Dyken, P., *Psychiatr Neurol Neurochir 10:*77-121 (1967)), this condition is believed to result from imbalance of neural transmission in the basal ganglia, since cases of secondary dystonia reveal lesions in the caudate nucleus, putamen, and globus pallidus, as well as the thalamus and rostral brain (Dooling, E. C., & Adams, R. D., *Brain 98:*29-48 (1975); Bhatia, K. P., & Marsden, C. D., *Brain 117:*859-876 (1994); Kulisevsky, J., *et al., Movement Disorder 8:*239-240 (1993)). Dystonia may result, in particular, from disruption of dopaminergic neurotransmission as genes defective in dopa-responsive dystonia encode proteins in the dopamine pathway (above); drugs that block dopaminergic transmission via the D2 receptor can elicit acute or tardive dystonic symptoms (Christian, C. D., & Paulson, G., *N Eng J Med 259:*828-830 (1958); Burke, R. E., *et al., Neurology 32:*1335-1346 (1982)), and abnormally low levels of dopaminergic metabolites have been noted in the cerebrospinal fluid of some dystonic patients (Tabaddor, K., *et al., Neurology 28:*1249-1253 (1978); Wolfson, L. I., *et al., Adv Neurol 50:*177-181 (1988); Brashear, A., *et al., Mov Disord 2:*151-156 (1996)). The clinical features of early onset dystonia reflect developmental and somatotopic patterns in the basal ganglia (Fig. 6). Neuromorphologic and physiologic studies in experimental animals have demonstrated an anatomic gradient of postnatal modelling in neural cell adhesion molecules (Szele, F.G., *et al., Neurosci 60:*133-144 (1994)) and dopaminergic innervation of the striatum (Graybiel, A. M., *Neuroscience 13:*1157-1187 (1984)). This developmental gradient overlies a somatotopic distribution of neurons in the basal ganglia subserving movements of different body parts (Crutcher, M. D., & DeLong, M. R., *Exp Brain Res 53:*233-243 (1984)), corresponding roughly to an inverted homunculus. In dystonia patients with the AJ founder mutation, the earlier the onset of symptoms between about 6 and 24 yrs of age, the more likely they are to occur in a lower limb and to generalize to other upper body parts (Bressman, S. B., *et al., Annal Neurol 36:*771-777 (1994b)). These ages appear to define a developmental period

of susceptibility; as carrier's of the same mutation who manifest no symptoms by 28 yrs will usually remain symptom-free for life. This developmental model of neuronal involvement in dystonia provides a platform for evaluating subtle neuromorphologic and physiologic changes in the basal ganglia. The ability to identify cells expressing the torsinA transcript should help to identify the neurons involved in dystonia.

The apparent lack of neuronal degeneration in early onset dystonia provides hope for therapeutic intervention. In fact, even delaying onset of symptoms in this form of dystonia might result in a milder phenotype. Most patients with dopa-responsive dystonia, a condition that mimics DYT1 clinically, are virtually cured by administration of levo-dopa and this treatment remains effective over the patient's lifetime. A minority of patients with typical early onset dystonia have shown improvement with various medications, including high dose anticholinergics, dopamine agonists and antagonists, and GABAergic agents, and in some cases drugs have been terminated without remission after the age of susceptibility is past (Fahn, S., *et al.*, in *Movement Disorders 2*, Marsden, C. D. & Fahn, S., eds., London: Butterworths (1987), pp. 332-358; Pranzatetti, M. R., *J Child Neurol 11*:355-369 (1996); Bressman, S. B., & Greene, P. E., *Neurol Clin 8*:51-75 (1990)). Further, patients with the best response to anticholinergics tend to be those who are treated early in the course of the disease (Greene, P., *et al.*, *Mov Disord 2*:237-254 (1988)).

* * * * *

All publications mentioned hereinabove are hereby incorporated in their entirety by reference.

While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be appreciated by one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention and appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT/INVENTOR: OZELIUS, LAURIE
BREAKEFIELD, XANDRA

(ii) TITLE OF INVENTION: TORSIN, TORSIN GENES, AND METHODS OF USE

(iii) NUMBER OF SEQUENCES: 8

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
(B) STREET: 1100 NEW YORK AVENUE, SUITE 600
(C) CITY: WASHINGTON
(D) STATE: DC
(E) COUNTRY: USA
(F) ZIP: 20005-3934

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: To be assigned
(B) FILING DATE: Herewith
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 60/050,244
(B) FILING DATE: 19-JUN-1997
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: GOLDSTEIN, JORGE A.
(B) REGISTRATION NUMBER: 29,021
(C) REFERENCE/DOCKET NUMBER: 0609.438PC01

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (202) 371-2600

(B) TELEFAX: (202) 371-2540

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2597 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 568..1563

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CTGAAAATAG CTNCTTATTA TTATTATTAT TATTATTATT ATTTGCGGGA GGGAGCACAG 60

TCTTGCTCTG TCTCCCAGGC TGGAGNTGCA GTGGTGAGAT CTCGGCTCAC TGCAATCTCC 120

GCCTCCTGGG TTCAAGNGCA GTTGCTCATG TGTCAGCCTC CCCAGTAGCT AGGGCTACAG 180

GTGCCTACCA CCACACCGGC TAATTTTATA TTTTTAGTAG AGACGTGGTT TCACCATGTT 240

GGTCAGGCTG GTCTCGAACT CCTGACCTCA GGTGATCCGC CCNCCTCAGC CTNCCCAAAG 300

GGCTGGGATT ACAGGCAGGA GCCACCATNC CTGGNAAAAA TAACGTCCAT AAACAAAAAC 360

ACGTGGCCAA CAGGGCGGAG CAGAACCGAG TTTCCGGAAG CAAAACAGGG CTTTGTACCG 420

AACAAAGATG GCGGCCGCCG GCGTCGGGAG GAGGGCTGCC CTGAAGAAAG ATGGCCTCCG 480

CGAGAGGAGG AANCCGGAAG CGTGGGTCTG GCGGCTGCAC CGGTTCGCGG TCGGCGCGAG 540

AACAAGCAGG GTGGCGCGGG TCCGGGC ATG AAG CTG GGC CGG GCC GTG CTG 591
Met Lys Leu Gly Arg Ala Val Leu
1 5

```
GGC CTG CTG CTG CTG GCG CCG TCC GTG GTG CAG GCG GTG GAG CCC ATC        639
Gly Leu Leu Leu Leu Ala Pro Ser Val Val Gln Ala Val Glu Pro Ile
     10                  15                  20

AGC CTG GGA CTG GCC CTG GCC GGC GTC CTC ACC GGC TAC ATC TAC CCG        687
Ser Leu Gly Leu Ala Leu Ala Gly Val Leu Thr Gly Tyr Ile Tyr Pro
 25                  30                  35                  40

CGT CTC TAC TGC CTC TTC GCC GAG TGC TGC GGG CAG AAG CGG AGC CTT        735
Arg Leu Tyr Cys Leu Phe Ala Glu Cys Cys Gly Gln Lys Arg Ser Leu
                 45                  50                  55

AGC CGG GAG GCA CTG CAG AAG GAT CTG GAC GAC AAC CTC TTT GGA CAG        783
Ser Arg Glu Ala Leu Gln Lys Asp Leu Asp Asp Asn Leu Phe Gly Gln
             60                  65                  70

CAT CTT GCA AAG AAA ATC ATC TTA AAT GCC GTG TTT GGT TTC ATA AAC        831
His Leu Ala Lys Lys Ile Ile Leu Asn Ala Val Phe Gly Phe Ile Asn
         75                  80                  85

AAC CCA AAG CCC AAG AAA CCT CTC ACG CTC TCC CTG CAC GGG TGG ACA        879
Asn Pro Lys Pro Lys Lys Pro Leu Thr Leu Ser Leu His Gly Trp Thr
     90                  95                  100

GGC ACC GGC AAA AAT TTC GTC AGC AAG ATC ATC GCA GAG AAT ATT TAC        927
Gly Thr Gly Lys Asn Phe Val Ser Lys Ile Ile Ala Glu Asn Ile Tyr
105                 110                 115                 120

GAG GGT GGT CTG AAC AGT GAC TAT GTC CAC CTG TTT GTG GCC ACA TTG        975
Glu Gly Gly Leu Asn Ser Asp Tyr Val His Leu Phe Val Ala Thr Leu
                125                 130                 135

CAC TTT CCA CAT GCT TCA AAC ATC ACC TTG TAC AAG GAT CAG TTA CAG       1023
His Phe Pro His Ala Ser Asn Ile Thr Leu Tyr Lys Asp Gln Leu Gln
            140                 145                 150

TTG TGG ATT CGA GGC AAC GTG AGT GCC TGT GCG AGG TCC ATC TTC ATA       1071
Leu Trp Ile Arg Gly Asn Val Ser Ala Cys Ala Arg Ser Ile Phe Ile
        155                 160                 165
```

```
TTT GAT GAA ATG GAT AAG ATG CAT GCA GGC CTC ATA GAT GCC ATC AAG      1119
Phe Asp Glu Met Asp Lys Met His Ala Gly Leu Ile Asp Ala Ile Lys
    170                 175                 180

CCT TTC CTC GAC TAT TAT GAC CTG GTG GAT GGG GTC TCC TAC CAG AAA      1167
Pro Phe Leu Asp Tyr Tyr Asp Leu Val Asp Gly Val Ser Tyr Gln Lys
185                 190                 195                 200

GCC ATG TTC ATA TTT CTC AGC AAT GCT GGA GCA GAA AGG ATC ACA GAT      1215
Ala Met Phe Ile Phe Leu Ser Asn Ala Gly Ala Glu Arg Ile Thr Asp
                205                 210                 215

GTG GCT TTG GAT TTC TGG AGG AGT GGA AAG CAG AGG GAA GAC ATC AAG      1263
Val Ala Leu Asp Phe Trp Arg Ser Gly Lys Gln Arg Glu Asp Ile Lys
            220                 225                 230

CTC AAA GAC ATT GAA CAC GCG TTG TCT GTG TCG GTT TTC AAT AAC AAG      1311
Leu Lys Asp Ile Glu His Ala Leu Ser Val Ser Val Phe Asn Asn Lys
        235                 240                 245

AAC AGT GGC TTC TGG CAC AGC AGC TTA ATT GAC CGG AAC CTC ATT GAT      1359
Asn Ser Gly Phe Trp His Ser Ser Leu Ile Asp Arg Asn Leu Ile Asp
    250                 255                 260

TAT TTT GTT CCC TTC CTC CCC CTG GAA TAC AAA CAC CTA AAA ATG TGT      1407
Tyr Phe Val Pro Phe Leu Pro Leu Glu Tyr Lys His Leu Lys Met Cys
265                 270                 275                 280

ATC CGA GTG GAA ATG CAG TCC CGA GGC TAT GAA ATT GAT GAA GAC ATT      1455
Ile Arg Val Glu Met Gln Ser Arg Gly Tyr Glu Ile Asp Glu Asp Ile
                285                 290                 295

GTA AGC AGA GTG GCT GAG GAG ATG ACA TTT TTC CCC AAA GAG GAG AGA      1503
Val Ser Arg Val Ala Glu Glu Met Thr Phe Phe Pro Lys Glu Glu Arg
            300                 305                 310

GTT TTC TCA GAT AAA GGC TGC AAA ACG GTG TTC ACC AAG TTA GAT TAT      1551
Val Phe Ser Asp Lys Gly Cys Lys Thr Val Phe Thr Lys Leu Asp Tyr
        315                 320                 325
```

TAC TAC GAT GAT TGACAGTCAT GATTGGCAGC CGGAGTCACT GCCTGGAGTT 1603
Tyr Tyr Asp Asp
330

GGAAAAGAAA CAACACTCAG TCCTTCCACA CTTCCACCCC CAGCTCCTTT CCCTGGAAGA 1663

GGAATCCAGT GAATGTTCCT GTTTGATGTG ACAGGAATTC TCCCTGGCAT TGTTTCCACC 1723

CCCTGGTGCC TGCAGGCCAC CCAGGGACCA CGGGCGAGGA CGTGAAGCCT CCCGAACACG 1783

CACAGAAGGA AGGAGCCAGC TCCCAGCCCA CTCATCGCAG GGCTCATGAT TTTTTACAAA 1843

TTATGTTTTA ATTCCAAGTG TTTCTGTTTC AAGGAAGGAT GAATAAGTTT TATTGAAAAT 1903

GTGGTAACTT TATTTAAAAT GATTTTTAAC ATTATGAGAG ACTGCTCAGA TTCTAAGTTG 1963

TTGGCCTTGT GTGTGTGTTT TTTTTTAAGT TCTCATCATT ATTACATAGA CTGTGAAGTA 2023

TCTTTACTGG AAATGAGCCC AAGCACACAT GCATGGCATT TGTTCCTGAA CAGGAGGGCA 2083

TCCCTGGGGA TGTGGCTGGA GCATGAGCCA GCTCTGTCCC AGGATGGTCC CAGCGGATGC 2143

TGCCAGGGGC AGTGAAGTGT TTAGGTGAAG GACAAGTAGG TAAGAGGACG CCTTCAGGCA 2203

CCACAGATAA GCCTGAAACA GCCTCTCCAA GGGTTTTCAC CTTAGCAACA ATGGGAGCTG 2263

TGGGAGTGAT TTTGGCCACA CTGTCAACAT TTGTTAGAAC CAGTCTTTTG AAAGAAAAGT 2323

ATTTCCAACT TGTCACTTGC CAGTCACTCC GTTTTGCAAA AGGTGGCCCT TCACTGTCCA 2383

TTCCAAATAG CCCACACGTG CTCTCTGCTG GATTCTAAAT TATGTGAATT TTGCCATATT 2443

AAATCTTCCT CATTTATACT ATTATTTGTT ACGTTCAATC AGAATCCCCG AAACCTCCTA 2503

TAAAGCTTAG CTGCCCCTTC TGAGGATGCT GAGAACGGTG TCTTTCTTTA TAAATGCAAA 2563

TGGCTACCGT TTTACAATAA AATTTTGCAT GTGC 2597

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 332 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Lys Leu Gly Arg Ala Val Leu Gly Leu Leu Leu Leu Ala Pro Ser
1 5 10 15

Val Val Gln Ala Val Glu Pro Ile Ser Leu Gly Leu Ala Leu Ala Gly
20 25 30

Val Leu Thr Gly Tyr Ile Tyr Pro Arg Leu Tyr Cys Leu Phe Ala Glu
35 40 45

Cys Cys Gly Gln Lys Arg Ser Leu Ser Arg Glu Ala Leu Gln Lys Asp
50 55 60

Leu Asp Asp Asn Leu Phe Gly Gln His Leu Ala Lys Lys Ile Ile Leu
65 70 75 80

Asn Ala Val Phe Gly Phe Ile Asn Asn Pro Lys Pro Lys Lys Pro Leu
85 90 95

Thr Leu Ser Leu His Gly Trp Thr Gly Thr Gly Lys Asn Phe Val Ser
100 105 110

Lys Ile Ile Ala Glu Asn Ile Tyr Glu Gly Gly Leu Asn Ser Asp Tyr
115 120 125

Val His Leu Phe Val Ala Thr Leu His Phe Pro His Ala Ser Asn Ile
130 135 140

Thr Leu Tyr Lys Asp Gln Leu Gln Leu Trp Ile Arg Gly Asn Val Ser
145 150 155 160

Ala Cys Ala Arg Ser Ile Phe Ile Phe Asp Glu Met Asp Lys Met His
165 170 175

Ala Gly Leu Ile Asp Ala Ile Lys Pro Phe Leu Asp Tyr Tyr Asp Leu
180 185 190

Val Asp Gly Val Ser Tyr Gln Lys Ala Met Phe Ile Phe Leu Ser Asn
195 200 205

Ala Gly Ala Glu Arg Ile Thr Asp Val Ala Leu Asp Phe Trp Arg Ser
210 215 220

Gly Lys Gln Arg Glu Asp Ile Lys Leu Lys Asp Ile Glu His Ala Leu
225 230 235 240

Ser Val Ser Val Phe Asn Asn Lys Asn Ser Gly Phe Trp His Ser Ser
245 250 255

Leu Ile Asp Arg Asn Leu Ile Asp Tyr Phe Val Pro Phe Leu Pro Leu
260 265 270

Glu Tyr Lys His Leu Lys Met Cys Ile Arg Val Glu Met Gln Ser Arg
275 280 285

Gly Tyr Glu Ile Asp Glu Asp Ile Val Ser Arg Val Ala Glu Glu Met
290 295 300

Thr Phe Phe Pro Lys Glu Glu Arg Val Phe Ser Asp Lys Gly Cys Lys
305 310 315 320

Thr Val Phe Thr Lys Leu Asp Tyr Tyr Tyr Asp Asp
325 330

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3568 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 994..1863

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TGATCCGCCT GCCTCGGCCT CCCAAAGTGC TGGGATTACA GGCGTGAGCG CCGCGCCCGG 60

CCAGCCTGAG ACAGTTTCGC TCTTGTCGCC CAGGCTGGAG TGCAGTGGCA CGATCTCGGC 120

TAACTGCAAC CTCCGCCTCC TGGGTTCAAG AGATTCTCCT GCCTCAACCT CCGAGTAGCT 180

GGGATTACAG GCGYSCGCCR CSMCRSCCAG CNTTTTTTTT TTTTTTTGAG ACAGTTTCGC 240

TCTTGTCGCC AGGCTGGAGT GCAGTGNNNG ANCTCGGCTA ACTGCAACTC CGCCTCCTGG 300

GTTCAAGAGA TTCTNCTGCC TCAACTCCCG AGTAGCTGGG ATTATAGGNG NCCGCNACCA 360

CACCATCTAA TTTTTTGTAT TTTTAGTAGA GACGGGGTTT CGCACGTTGA GCAGGCTGGT 420

CTCGAACTCC TGACATCAGG TGATCCGCCC GATTCAGCTT CCCAAAGTGC TGGGATTACA 480

GGCGTGAGCA CGGCGCCCGG CCAAAAAAAA AATATTTTTT TTTTTTTTTT AGATATTTTT 540

TCACTCTTGT TGCCCAGGCT GGAGTGAAAT GGCGTGATCT CGGCTCGGCC TCCCAAAGTC 600

CTGGGATTAC AGGCGTGACA CCGNGCCCGG SCCGAAAAAA TWTTTTTAAA AGAAAAAGGG 660

AAACAAAMAG TCTCCTACAC CTTCGRCCAC TCCCAAGAAC GATGGSACSS CCTCCTCTYC 720

GNCCCTMACC AACCATGGCC GNCCCNAAGG GAGTGGGGCG GGTCTGCGGG GCGGAAGTGA 780

CGSACGAGAG GAAGTCCGTC CTGCGCTTGG CCGCGGGGCG CCTGGCTCAG TGGCTTCTGC 840

GGGCTTCGAG GAGCGGGATG TTGCGGGCTG GGTGGCTCCG GGCGCGGCGG CGCTGGCGCT 900

GCTGCTGGCG GCCCGAGTGG TGGCGGAGTT CGAGCCCATC ACCGTGGGCC TAGCCATCGG 960

```
GCAGGTCGGC CATCACCGGC TACCTGTCCT ACA ATG ACA TCT ACT GCC CGT TCG      1014
                                     Met Thr Ser Thr Ala Arg Ser
                                     335

CCG AGT GCT GCC GGC GAG GAG CGG CCG CTC AAC GCT TCG GCT CTC AAG      1062
Pro Ser Ala Ala Gly Glu Glu Arg Pro Leu Asn Ala Ser Ala Leu Lys
340                 345                 350                 355

CTG GAT TTG GAG GAG AAG CTG TTT GGA CAG CAT CTA GCC ACG GAA GTG      1110
Leu Asp Leu Glu Glu Lys Leu Phe Gly Gln His Leu Ala Thr Glu Val
                360                 365                 370

ATT TTC AAG GCG CTG ACT GGC TTC AGG AAC AAC AAA AAT CCC AAG AAA      1158
Ile Phe Lys Ala Leu Thr Gly Phe Arg Asn Asn Lys Asn Pro Lys Lys
            375                 380                 385

CCA CTG ACC CTT TCC TTA CAC GGC TGG GCT GGC ACA GGC AAG AAT TTT      1206
Pro Leu Thr Leu Ser Leu His Gly Trp Ala Gly Thr Gly Lys Asn Phe
        390                 395                 400

GTC AGT CAA ATT GTG GCT GAA AAT CTT CAC CCA AAA GGT CTG AAG AGT      1254
Val Ser Gln Ile Val Ala Glu Asn Leu His Pro Lys Gly Leu Lys Ser
    405                 410                 415

AAC TTT GTC CAC CTG TTT GTA TCG ACT CTG CAC TTC CCT CAT GAG CAG      1302
Asn Phe Val His Leu Phe Val Ser Thr Leu His Phe Pro His Glu Gln
420                 425                 430                 435

AAG ATA AAA CTG TAC CAG GAC CAG TTA CAG AAG TGG ATC CGC GGT AAT      1350
Lys Ile Lys Leu Tyr Gln Asp Gln Leu Gln Lys Trp Ile Arg Gly Asn
                440                 445                 450

GTG AGT GCA TGT GCG AAC TCT GTT TTC ATA TTT GAC GAG ATG GAT AAA      1398
Val Ser Ala Cys Ala Asn Ser Val Phe Ile Phe Asp Glu Met Asp Lys
            455                 460                 465

TTG CAC CCC GGG ATC ATT GAC GCA ATC AAG CCG TTT CTA GAC TAC TAC      1446
Leu His Pro Gly Ile Ile Asp Ala Ile Lys Pro Phe Leu Asp Tyr Tyr
        470                 475                 480
```

```
GAG CAG GTT GAC GGA GTG TCT TAC CGC AAA GCC ATC TTC ATC TTT CTC      1494
Glu Gln Val Asp Gly Val Ser Tyr Arg Lys Ala Ile Phe Ile Phe Leu
    485                 490                 495

AGC AAT GCA GGC GGG GAC CTT ATA ACT AAG ACG GCT CTT GAC TTT TGG      1542
Ser Asn Ala Gly Gly Asp Leu Ile Thr Lys Thr Ala Leu Asp Phe Trp
500                 505                 510                 515

CGG GCC GGA AGA AAG AGG GAA GAC ATT CAG CTG AAG GAC CTG GAA CCT      1590
Arg Ala Gly Arg Lys Arg Glu Asp Ile Gln Leu Lys Asp Leu Glu Pro
                520                 525                 530

GTA CTG TCT GTC GGA GTC TTC AAT AAT AAA CAC AGT GGC CTG TGG CAC      1638
Val Leu Ser Val Gly Val Phe Asn Asn Lys His Ser Gly Leu Trp His
            535                 540                 545

AGT GGA CTG ATC GAC AAA AAC CTC ATT GAT TAC TTT ATC CCC TTC CTG      1686
Ser Gly Leu Ile Asp Lys Asn Leu Ile Asp Tyr Phe Ile Pro Phe Leu
        550                 555                 560

CCT TTG GAG TAC AGA CAT GTG AAA ATG TGT GTG AGG GCC GAG ATG AGG      1734
Pro Leu Glu Tyr Arg His Val Lys Met Cys Val Arg Ala Glu Met Arg
    565                 570                 575

GCC CGT GGT TCT GCC ATA GAT GAA GAC ATT GTC ACA AGA GTG GCA GAG      1782
Ala Arg Gly Ser Ala Ile Asp Glu Asp Ile Val Thr Arg Val Ala Glu
580                 585                 590                 595

GAA ATG ACG TTT TTC CCC AGA GAC GAG AAA ATC TAC TCA GAC AAG GGC      1830
Glu Met Thr Phe Phe Pro Arg Asp Glu Lys Ile Tyr Ser Asp Lys Gly
                600                 605                 610

TGC AAG ACT GTG CAG TCG CGG CTG GAT TTC CAC TGAGCTCCTA TCCAGATGGG    1883
Cys Lys Thr Val Gln Ser Arg Leu Asp Phe His
            615                 620

GTAGGAGACA GCTGGGAGGC TCCGCACGCC AGAGGCCTTG CCTTTCAGAA GAACCCTGAA    1943

GACCGCTTTG GGGTTTTGCC TGTTTGCACC TTAGACTTTT GGGTATAGAA TCTTTTTTTT    2003
```

GAGAAGAGGT CTCACTCCGT CATCCAAGCT GGAGTGCAGT GGTGCAATCC TCAACTCACT 2063

GCAACCTCCG CTCCCGGTTT GAGTGATTCT CATGCCTCAG CCTCCCGAGT AGCTGGGATT 2123

ACAGGCATGA GCCACTGTGC CCAGCTGGGA TATAGAATCT AAGAGTTGAT TGTGGAAAAC 2183

ACGTGAATCT ATTGCGCGCA TTTGTCATTT AGCAAGATGG CAGCAGTCCA GCTGTTCTTT 2243

GCAGCTGGAG ATGAACTTTT AAAAATCCCC TTCACACTTA ATGTACTGAC CGAGACAGAA 2303

GTACCTGAAA ACAGCTGTGC ATGGCAGGCC CGGCAATAGC TTCTGACCCA CAGCACCCGC 2363

GCCTCAGAAG CTACGGTCAC AACTAAAGGA GTCCAGGGAC TTGCTGCAGG CTGGGGGGCA 2423

CTGGGTGGTT CTCACCAGCA GGCTGCGGGG CACTGTGTTC TCATTGGCCA AAAACATCCT 2483

TTTGCTCTGT CTCGTTCTTT ACACAGAGTT CACTGACTTG AAGTATACTC AGTTAAAATC 2543

GGGGCTGGAG GTGCAGACGG TGTCTGACCG GAGGATGTGG CCGTGCCCGC CGAGCACTCT 2603

TGATCTGAGC TGACCTGTGT GTGTGTGTGG GGGGGGGTGG GGCCTTCACC TAAGACCTCT 2663

GCAGCAGACC TGGACAGACA GGCCCCTCCC GCCTGTCCAT CGCTCTAGCT GCTAATACAG 2723

CCCTGGCTGT GGAATCCTTC ACCGTCTCAG CTGGTATCAG CCCCAGCCTG CCTTGGTGCC 2783

ATATCTCAGC TTGGATCTCT GCTAGAGTCC CCCCAACCAT ATATCATAGA GTTGAATCAC 2843

AATGAGACCG TTGGCTTTGA ATTTGAGTCG TTGGTTCCCA TGGTGAGATG CTTGTTAAGA 2903

CTTTATACTT GGGTCAATCT CTCACTTTAT TTTGTAGAAC CATTTGAAAT CCTAGGATGT 2963

GCTTGTTCTG GAAGGATGAC ATGGGCCCAG ACTGAACAAG TCAGCTTGAT GATCTTAAAT 3023

GATGGAAGTA TAGGACGTTG CTTATTTTAA AACAAGGGAA GGACACAAAA TGGAATGACT 3083

GCCTTAGTCC TTTCTCAGAT ACTCCTTAAA ACAATTTTTT ATTGTTTAAA TTTGTGGTAA 3143

TACATGGTCA CAACCGTGGA TCAAACAAGG TCAGTCTAAA GTGGCAGGTC CTAGGTGTGA 3203

CCTGATACCA CCACCCTTTG TGGCAGCACC GGGCTGGACT GCCCTGATCC CTGGGACGTG 3263

AGACTTAGCT TCCAGCCAGT GTGAATCATT GTATCTGTCT CATAATCACA GCACAGCTGC 3323

AGACACAACA ACGTGCAGCA TTTTTTACAT AAAAATATGG TAGAATTAAT TTATGACATG 3383

GAAATGCCTT ACGTGGTATC ACACTTAGTC TTGAAAAAAA CACCAAGGTG ACGTTTAAAA 3443

TTTTTAGTAC ATATCCTCAA ATTGGAGCTA AGTTATACTT CTTTTATAAC CTTTTGGGCA 3503

TCTGGTCGAG AGAAGACAAG ATTTTCTCTA TTTACAGTGA GGCAATAAAT ATGTTTGCCA 3563

CCTTT 3568

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 290 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Thr Ser Thr Ala Arg Ser Pro Ser Ala Ala Gly Glu Glu Arg Pro
1 5 10 15

Leu Asn Ala Ser Ala Leu Lys Leu Asp Leu Glu Glu Lys Leu Phe Gly
20 25 30

Gln His Leu Ala Thr Glu Val Ile Phe Lys Ala Leu Thr Gly Phe Arg
35 40 45

Asn Asn Lys Asn Pro Lys Lys Pro Leu Thr Leu Ser Leu His Gly Trp
50 55 60

Ala Gly Thr Gly Lys Asn Phe Val Ser Gln Ile Val Ala Glu Asn Leu
65 70 75 80

His Pro Lys Gly Leu Lys Ser Asn Phe Val His Leu Phe Val Ser Thr
85 90 95

Leu His Phe Pro His Glu Gln Lys Ile Lys Leu Tyr Gln Asp Gln Leu
100 105 110

Gln Lys Trp Ile Arg Gly Asn Val Ser Ala Cys Ala Asn Ser Val Phe
115 120 125

Ile Phe Asp Glu Met Asp Lys Leu His Pro Gly Ile Ile Asp Ala Ile
130 135 140

Lys Pro Phe Leu Asp Tyr Tyr Glu Gln Val Asp Gly Val Ser Tyr Arg
145 150 155 160

Lys Ala Ile Phe Ile Phe Leu Ser Asn Ala Gly Gly Asp Leu Ile Thr
165 170 175

Lys Thr Ala Leu Asp Phe Trp Arg Ala Gly Arg Lys Arg Glu Asp Ile
180 185 190

Gln Leu Lys Asp Leu Glu Pro Val Leu Ser Val Gly Val Phe Asn Asn
195 200 205

Lys His Ser Gly Leu Trp His Ser Gly Leu Ile Asp Lys Asn Leu Ile
210 215 220

Asp Tyr Phe Ile Pro Phe Leu Pro Leu Glu Tyr Arg His Val Lys Met
225 230 235 240

Cys Val Arg Ala Glu Met Arg Ala Arg Gly Ser Ala Ile Asp Glu Asp
245 250 255

Ile Val Thr Arg Val Ala Glu Glu Met Thr Phe Phe Pro Arg Asp Glu
260 265 270

Lys Ile Tyr Ser Asp Lys Gly Cys Lys Thr Val Gln Ser Arg Leu Asp
275 280 285

Phe His

290

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2072 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 43..1038

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
CGCGGTCGGC GCGAGAACAA GCAGGGTGGC GCGGGTCCGG GC ATG AAG CTG GGC          54
                                                 Met Lys Leu Gly

CGG GCC GTG CTG GGC CTG CTG CTG CTG GCG CCG TCC GTG GTG CAG GCG          102
Arg Ala Val Leu Gly Leu Leu Leu Leu Ala Pro Ser Val Val Gln Ala
295                 300                 305                 310

GTG GAG CCC ATC AGC CTG GGA CTG GCC CTG GCC GGC GTC CTC ACC GGC          150
Val Glu Pro Ile Ser Leu Gly Leu Ala Leu Ala Gly Val Leu Thr Gly
                315                 320                 325

TAC ATC TAC CCG CGT CTC TAC TGC CTC TTC GCC GAG TGC TGC GGG CAG          198
Tyr Ile Tyr Pro Arg Leu Tyr Cys Leu Phe Ala Glu Cys Cys Gly Gln
            330                 335                 340

AAG CGG AGC CTT AGC CGG GAG GCA CTG CAG AAG GAT CTG GAC GAC AAC          246
Lys Arg Ser Leu Ser Arg Glu Ala Leu Gln Lys Asp Leu Asp Asp Asn
        345                 350                 355

CTC TTT GGA CAG CAT CTT GCA AAG AAA ATC ATC TTA AAT GCC GTG TTT          294
Leu Phe Gly Gln His Leu Ala Lys Lys Ile Ile Leu Asn Ala Val Phe
```

360 365 370

GGT TTC ATA AAC AAC CCA AAG CCC AAG AAA CCT CTC ACG CTC TCC CTG 342
Gly Phe Ile Asn Asn Pro Lys Pro Lys Lys Pro Leu Thr Leu Ser Leu
375 380 385 390

CAC GGG TGG ACA GGC ACC GGC AAA AAT TTC GTC AGC AAG ATC ATC GCA 390
His Gly Trp Thr Gly Thr Gly Lys Asn Phe Val Ser Lys Ile Ile Ala
395 400 405

GAG AAT ATT TAC GAG GGT GGT CTG AAC AGT GAC TAT GTC CAC CTG TTT 438
Glu Asn Ile Tyr Glu Gly Gly Leu Asn Ser Asp Tyr Val His Leu Phe
410 415 420

GTG GCC ACA TTG CAC TTT CCA CAT GCT TCA AAC ATC ACC TTG TAC AAG 486
Val Ala Thr Leu His Phe Pro His Ala Ser Asn Ile Thr Leu Tyr Lys
425 430 435

GAT CAG TTA CAG TTG TGG ATT CGA GGC AAC GTG AGT GCC TGT GCG AGG 534
Asp Gln Leu Gln Leu Trp Ile Arg Gly Asn Val Ser Ala Cys Ala Arg
440 445 450

TCC ATC TTC ATA TTT GAT GAA ATG GAT AAG ATG CAT GCA GGC CTC ATA 582
Ser Ile Phe Ile Phe Asp Glu Met Asp Lys Met His Ala Gly Leu Ile
455 460 465 470

GAT GCC ATC AAG CCT TTC CTC GAC TAT TAT GAC CTG GTG GAT GGG GTC 630
Asp Ala Ile Lys Pro Phe Leu Asp Tyr Tyr Asp Leu Val Asp Gly Val
475 480 485

TCC TAC CAG AAA GCC ATG TTC ATA TTT CTC AGC AAT GCT GGA GCA GAA 678
Ser Tyr Gln Lys Ala Met Phe Ile Phe Leu Ser Asn Ala Gly Ala Glu
490 495 500

AGG ATC ACA GAT GTG GCT TTG GAT TTC TGG AGG AGT GGA AAG CAG AGG 726
Arg Ile Thr Asp Val Ala Leu Asp Phe Trp Arg Ser Gly Lys Gln Arg
505 510 515

GAA GAC ATC AAG CTC AAA GAC ATT GAA CAC GCG TTG TCT GTG TCG GTT 774
Glu Asp Ile Lys Leu Lys Asp Ile Glu His Ala Leu Ser Val Ser Val

```
520                 525                 530

TTC AAT AAC AAG AAC AGT GGC TTC TGG CAC AGC AGC TTA ATT GAC CGG      822
Phe Asn Asn Lys Asn Ser Gly Phe Trp His Ser Ser Leu Ile Asp Arg
535                 540                 545                 550

AAC CTC ATT GAT TAT TTT GTT CCC TTC CTC CCC CTG GAA TAC AAA CAC      870
Asn Leu Ile Asp Tyr Phe Val Pro Phe Leu Pro Leu Glu Tyr Lys His
                555                 560                 565

CTA AAA ATG TGT ATC CGA GTG GAA ATG CAG TCC CGA GGC TAT GAA ATT      918
Leu Lys Met Cys Ile Arg Val Glu Met Gln Ser Arg Gly Tyr Glu Ile
            570                 575                 580

GAT GAA GAC ATT GTA AGC AGA GTG GCT GAG GAG ATG ACA TTT TTC CCC      966
Asp Glu Asp Ile Val Ser Arg Val Ala Glu Glu Met Thr Phe Phe Pro
        585                 590                 595

AAA GAG GAG AGA GTT TTC TCA GAT AAA GGC TGC AAA ACG GTG TTC ACC     1014
Lys Glu Glu Arg Val Phe Ser Asp Lys Gly Cys Lys Thr Val Phe Thr
    600                 605                 610

AAG TTA GAT TAT TAC TAC GAT GAT TGACAGTCAT GATTGGCAGC CGGAGTCACT     1068
Lys Leu Asp Tyr Tyr Tyr Asp Asp
615                 620

GCCTGGAGTT GGAAAAGAAA CAACACTCAG TCCTTCCACA CTTCCACCCC CAGCTCCTTT    1128

CCCTGGAAGA GGAATCCAGT GAATGTTCCT GTTTGATGTG ACAGGAATTC TCCCTGGCAT    1188

TGTTTCCACC CCCTGGTGCC TGCAGGCCAC CCAGGGACCA CGGGCGAGGA CGTGAAGCCT    1248

CCCGAACACG CACAGAAGGA AGGAGCCAGC TCCCAGCCCA CTCATCGCAG GGCTCATGAT    1308

TTTTTACAAA TTATGTTTTA ATTCCAAGTG TTTCTGTTTC AAGGAAGGAT GAATAAGTTT    1368

TATTGAAAAT GTGGTAACTT TATTTAAAAT GATTTTTAAC ATTATGAGAG ACTGCTCAGA    1428

TTCTAAGTTG TTGGCCTTGT GTGTGTGTTT TTTTTTAAGT TCTCATCATT ATTACATAGA    1488
```

CTGTGAAGTA TCTTTACTGG AAATGAGCCC AAGCACACAT GCATGGCATT TGTTCCTGAA 1548

CAGGAGGGCA TCCCTGGGGA TGTGGCTGGA GCATGAGCCA GCTCTGTCCC AGGATGGTCC 1608

CAGCGGATGC TGCCAGGGGC AGTGAAGTGT TTAGGTGAAG GACAAGTAGG TAAGAGGACG 1668

CCTTCAGGCA CCACAGATAA GCCTGAAACA GCCTCTCCAA GGGTTTTCAC CTTAGCAACA 1728

ATGGGAGCTG TGGGAGTGAT TTTGGCCACA CTGTCAACAT TTGTTAGAAC CAGTCTTTTG 1788

AAAGAAAAGT ATTTCCAACT TGTCACTTGC CAGTCACTCC GTTTTGCAAA AGGTGGCCCT 1848

TCACTGTCCA TTCCAAATAG CCCACACGTG CTCTCTGCTG GATTCTAAAT TATGTGAATT 1908

TTGCCATATT AAATCTTCCT CATTTATACT ATTATTTGTT ACGTTCAATC AGAATCCCCG 1968

AAACCTCCTA TAAAGCTTAG CTGCCCCTTC TGAGGATGCT GAGAACGGTG TCTTTCTTTA 2028

TAAATGCAAA TGGCTACCGT TTTACAATAA AATTTTGCAT GTGC 2072

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 332 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Lys Leu Gly Arg Ala Val Leu Gly Leu Leu Leu Leu Ala Pro Ser
1 5 10 15

Val Val Gln Ala Val Glu Pro Ile Ser Leu Gly Leu Ala Leu Ala Gly
20 25 30

Val Leu Thr Gly Tyr Ile Tyr Pro Arg Leu Tyr Cys Leu Phe Ala Glu
35 40 45

Cys Cys Gly Gln Lys Arg Ser Leu Ser Arg Glu Ala Leu Gln Lys Asp
50 55 60

Leu Asp Asp Asn Leu Phe Gly Gln His Leu Ala Lys Lys Ile Ile Leu
65 70 75 80

Asn Ala Val Phe Gly Phe Ile Asn Asn Pro Lys Pro Lys Lys Pro Leu
85 90 95

Thr Leu Ser Leu His Gly Trp Thr Gly Thr Gly Lys Asn Phe Val Ser
100 105 110

Lys Ile Ile Ala Glu Asn Ile Tyr Glu Gly Gly Leu Asn Ser Asp Tyr
115 120 125

Val His Leu Phe Val Ala Thr Leu His Phe Pro His Ala Ser Asn Ile
130 135 140

Thr Leu Tyr Lys Asp Gln Leu Gln Leu Trp Ile Arg Gly Asn Val Ser
145 150 155 160

Ala Cys Ala Arg Ser Ile Phe Ile Phe Asp Glu Met Asp Lys Met His
165 170 175

Ala Gly Leu Ile Asp Ala Ile Lys Pro Phe Leu Asp Tyr Tyr Asp Leu
180 185 190

Val Asp Gly Val Ser Tyr Gln Lys Ala Met Phe Ile Phe Leu Ser Asn
195 200 205

Ala Gly Ala Glu Arg Ile Thr Asp Val Ala Leu Asp Phe Trp Arg Ser
210 215 220

Gly Lys Gln Arg Glu Asp Ile Lys Leu Lys Asp Ile Glu His Ala Leu
225 230 235 240

Ser Val Ser Val Phe Asn Asn Lys Asn Ser Gly Phe Trp His Ser Ser
245 250 255

Leu Ile Asp Arg Asn Leu Ile Asp Tyr Phe Val Pro Phe Leu Pro Leu

260 265 270

Glu Tyr Lys His Leu Lys Met Cys Ile Arg Val Glu Met Gln Ser Arg
275 280 285

Gly Tyr Glu Ile Asp Glu Asp Ile Val Ser Arg Val Ala Glu Glu Met
290 295 300

Thr Phe Phe Pro Lys Glu Glu Arg Val Phe Ser Asp Lys Gly Cys Lys
305 310 315 320

Thr Val Phe Thr Lys Leu Asp Tyr Tyr Tyr Asp Asp
325 330

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2504 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..799

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

G GAT TTG GAG GAG AAG CTG TTT GGA CAG CAT CTA GCC ACG GAA GTG 46
Asp Leu Glu Glu Lys Leu Phe Gly Gln His Leu Ala Thr Glu Val
335 340 345

ATT TTC AAG GCG CTG ACT GGC TTC AGG AAC AAC AAA AAT CCC AAG AAA 94
Ile Phe Lys Ala Leu Thr Gly Phe Arg Asn Asn Lys Asn Pro Lys Lys
350 355 360

```
CCA CTG ACC CTT TCC TTA CAC GGC TGG GCT GGC ACA GGC AAG AAT TTT      142
Pro Leu Thr Leu Ser Leu His Gly Trp Ala Gly Thr Gly Lys Asn Phe
    365                 370                 375

GTC AGT CAA ATT GTG GCT GAA AAT CTT CAC CCA AAA GGT CTG AAG AGT      190
Val Ser Gln Ile Val Ala Glu Asn Leu His Pro Lys Gly Leu Lys Ser
380                 385                 390                 395

AAC TTT GTC CAC CTG TTT GTA TCG ACT CTG CAC TTC CCT CAT GAG CAG      238
Asn Phe Val His Leu Phe Val Ser Thr Leu His Phe Pro His Glu Gln
                400                 405                 410

AAG ATA AAA CTG TAC CAG GAC CAG TTA CAG AAG TGG ATC CGC GGT AAT      286
Lys Ile Lys Leu Tyr Gln Asp Gln Leu Gln Lys Trp Ile Arg Gly Asn
            415                 420                 425

GTG AGT GCA TGT GCG AAC TCT GTT TTC ATA TTT GAC GAG ATG GAT AAA      334
Val Ser Ala Cys Ala Asn Ser Val Phe Ile Phe Asp Glu Met Asp Lys
        430                 435                 440

TTG CAC CCC GGG ATC ATT GAC GCA ATC AAG CCG TTT CTA GAC TAC TAC      382
Leu His Pro Gly Ile Ile Asp Ala Ile Lys Pro Phe Leu Asp Tyr Tyr
    445                 450                 455

GAG CAG GTT GAC GGA GTG TCT TAC CGC AAA GCC ATC TTC ATC TTT CTC      430
Glu Gln Val Asp Gly Val Ser Tyr Arg Lys Ala Ile Phe Ile Phe Leu
460                 465                 470                 475

AGC AAT GCA GGC GGG GAC CTT ATA ACT AAG ACG GCT CTT GAC TTT TGG      478
Ser Asn Ala Gly Gly Asp Leu Ile Thr Lys Thr Ala Leu Asp Phe Trp
                480                 485                 490

CGG GCC GGA AGA AAG AGG GAA GAC ATT CAG CTG AAG GAC CTG GAA CCT      526
Arg Ala Gly Arg Lys Arg Glu Asp Ile Gln Leu Lys Asp Leu Glu Pro
            495                 500                 505

GTA CTG TCT GTC GGA GTC TTC AAT AAT AAA CAC AGT GGC CTG TGG CAC      574
Val Leu Ser Val Gly Val Phe Asn Asn Lys His Ser Gly Leu Trp His
        510                 515                 520
```

AGT GGA CTG ATC GAC AAA AAC CTC ATT GAT TAC TTT ATC CCC TTC CTG 622
Ser Gly Leu Ile Asp Lys Asn Leu Ile Asp Tyr Phe Ile Pro Phe Leu
525 530 535

CCT TTG GAG TAC AGA CAT GTG AAA ATG TGT GTG AGG GCC GAG ATG AGG 670
Pro Leu Glu Tyr Arg His Val Lys Met Cys Val Arg Ala Glu Met Arg
540 545 550 555

GCC CGT GGT TCT GCC ATA GAT GAA GAC ATT GTC ACA AGA GTG GCA GAG 718
Ala Arg Gly Ser Ala Ile Asp Glu Asp Ile Val Thr Arg Val Ala Glu
560 565 570

GAA ATG ACG TTT TTC CCC AGA GAC GAG AAA ATC TAC TCA GAC AAG GGC 766
Glu Met Thr Phe Phe Pro Arg Asp Glu Lys Ile Tyr Ser Asp Lys Gly
575 580 585

TGC AAG ACT GTG CAG TCG CGG CTG GAT TTC CAC TGAGCTCCTA TCCAGATGGG 819
Cys Lys Thr Val Gln Ser Arg Leu Asp Phe His
590 595

GTAGGAGACA GCTGGGAGGC TCCGCACGCC AGAGGCCTTG CCTTTCAGAA GAACCCTGAA 879

GACCGCTTTG GGGTTTTGCC TGTTTGCACC TTAGACTTTT GGGTATAGAA TCTTTTTTTT 939

GAGAAGAGGT CTCACTCCGT CATCCAAGCT GGAGTGCAGT GGTGCAATCC TCAACTCACT 999

GCAACCTCCG CTCCCGGTTT GAGTGATTCT CATGCCTCAG CCTCCCGAGT AGCTGGGATT 1059

ACAGGCATGA GCCACTGTGC CCAGCTGGGA TATAGAATCT AAGAGTTGAT TGTGGAAAAC 1119

ACGTGAATCT ATTGCGCGCA TTTGTCATTT AGCAAGATGG CAGCAGTCCA GCTGTTCTTT 1179

GCAGCTGGAG ATGAACTTTT AAAAATCCCC TTCACACTTA ATGTACTGAC CGAGACAGAA 1239

GTACCTGAAA ACAGCTGTGC ATGGCAGGCC CGGCAATAGC TTCTGACCCA CAGCACCCGC 1299

GCCTCAGAAG CTACGGTCAC AACTAAAGGA GTCCAGGGAC TTGCTGCAGG CTGGGGGGCA 1359

CTGGGTGGTT CTCACCAGCA GGCTGCGGGG CACTGTGTTC TCATTGGCCA AAAACATCCT 1419

TTTGCTCTGT CTCGTTCTTT ACACAGAGTT CACTGACTTG AAGTATACTC AGTTAAAATC 1479

GGGGCTGGAG GTGCAGACGG TGTCTGACCG GAGGATGTGG CCGTGCCCGC CGAGCACTCT 1539

TGATCTGAGC TGACCTGTGT GTGTGTGTGG GGGGGGGTGG GGCCTTCACC TAAGACCTCT 1599

GCAGCAGACC TGGACAGACA GGCCCCTCCC GCCTGTCCAT CGCTCTAGCT GCTAATACAG 1659

CCCTGGCTGT GGAATCCTTC ACCGTCTCAG CTGGTATCAG CCCCAGCCTG CCTTGGTGCC 1719

ATATCTCAGC TTGGATCTCT GCTAGAGTCC CCCCAACCAT ATATCATAGA GTTGAATCAC 1779

AATGAGACCG TTGGCTTTGA ATTTGAGTCG TTGGTTCCCA TGGTGAGATG CTTGTTAAGA 1839

CTTTATACTT GGGTCAATCT CTCACTTTAT TTTGTAGAAC CATTTGAAAT CCTAGGATGT 1899

GCTTGTTCTG GAAGGATGAC ATGGGCCCAG ACTGAACAAG TCAGCTTGAT GATCTTAAAT 1959

GATGGAAGTA TAGGACGTTG CTTATTTTAA AACAAGGGAA GGACACAAAA TGGAATGACT 2019

GCCTTAGTCC TTTCTCAGAT ACTCCTTAAA ACAATTTTTT ATTGTTTAAA TTTGTGGTAA 2079

TACATGGTCA CAACCGTGGA TCAAACAAGG TCAGTCTAAA GTGGCAGGTC CTAGGTGTGA 2139

CCTGATACCA CCACCCTTTG TGGCAGCACC GGGCTGGACT GCCCTGATCC CTGGGACGTG 2199

AGACTTAGCT TCCAGCCAGT GTGAATCATT GTATCTGTCT CATAATCACA GCACAGCTGC 2259

AGACACAACA ACGTGCAGCA TTTTTTACAT AAAAATATGG TAGAATTAAT TTATGACATG 2319

GAAATGCCTT ACGTGGTATC ACACTTAGTC TTGAAAAAAA CACCAAGGTG ACGTTTAAAA 2379

TTTTTAGTAC ATATCCTCAA ATTGGAGCTA AGTTATACTT CTTTTATAAC CTTTTGGGCA 2439

TCTGGTCGAG AGAAGACAAG ATTTTCTCTA TTTACAGTGA GGCAATAAAT ATGTTTGCCA 2499

CCTTT 2504

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 266 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Asp Leu Glu Glu Lys Leu Phe Gly Gln His Leu Ala Thr Glu Val Ile
 1               5                  10                  15

Phe Lys Ala Leu Thr Gly Phe Arg Asn Asn Lys Asn Pro Lys Lys Pro
            20                  25                  30

Leu Thr Leu Ser Leu His Gly Trp Ala Gly Thr Gly Lys Asn Phe Val
        35                  40                  45

Ser Gln Ile Val Ala Glu Asn Leu His Pro Lys Gly Leu Lys Ser Asn
    50                  55                  60

Phe Val His Leu Phe Val Ser Thr Leu His Phe Pro His Glu Gln Lys
65                  70                  75                  80

Ile Lys Leu Tyr Gln Asp Gln Leu Gln Lys Trp Ile Arg Gly Asn Val
                85                  90                  95

Ser Ala Cys Ala Asn Ser Val Phe Ile Phe Asp Glu Met Asp Lys Leu
            100                 105                 110

His Pro Gly Ile Ile Asp Ala Ile Lys Pro Phe Leu Asp Tyr Tyr Glu
        115                 120                 125

Gln Val Asp Gly Val Ser Tyr Arg Lys Ala Ile Phe Ile Phe Leu Ser
    130                 135                 140

Asn Ala Gly Gly Asp Leu Ile Thr Lys Thr Ala Leu Asp Phe Trp Arg
145                 150                 155                 160
```

Ala Gly Arg Lys Arg Glu Asp Ile Gln Leu Lys Asp Leu Glu Pro Val
165 170 175

Leu Ser Val Gly Val Phe Asn Asn Lys His Ser Gly Leu Trp His Ser
180 185 190

Gly Leu Ile Asp Lys Asn Leu Ile Asp Tyr Phe Ile Pro Phe Leu Pro
195 200 205

Leu Glu Tyr Arg His Val Lys Met Cys Val Arg Ala Glu Met Arg Ala
210 215 220

Arg Gly Ser Ala Ile Asp Glu Asp Ile Val Thr Arg Val Ala Glu Glu
225 230 235 240

Met Thr Phe Phe Pro Arg Asp Glu Lys Ile Tyr Ser Asp Lys Gly Cys
245 250 255

Lys Thr Val Gln Ser Arg Leu Asp Phe His
260 265

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM
(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page ____16_____, line ____14_____.

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☒

Name of depositary institution

American Type Culture Collection

Address of depositary institution *(including postal code and country)*

12301 Parklawn Drive
Rockville, MD 20852
United States of America

Now located at:
10801 University Blvd.
Manassas, VA 20110-2209
United States of America

Date of deposit	Accession Number
June 12, 1997	98454

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

Escherichia coli with cDNA in plasmid H4. DYT1

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications, e.g. "Accession Number of Deposit")*.

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Sonya Barnes PCT International Division	

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM
(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page ____16_____, line ____14_____.

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution

American Type Culture Collection

Address of depositary institution *(including postal code and country)*

12301 Parklawn Drive
Rockville, MD 20852
United States of America

Now located at:
10801 University Blvd.
Manassas, VA 20110-2209
United States of America

Date of deposit	Accession Number
June 12, 1997	98455

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

Escherichia coli containing human cDNA in plasmid 17

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the international Bureau later *(specify the general nature of the indications, e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application Sonya Barnes PCT International Division	☐ This sheet was received by the International Bureau on:

What Is Claimed Is:

1. An isolated nucleic acid molecule comprising a polynucleotide sequence at least 90% identical to a sequence selected from the group consisting of:

(a) a nucleotide sequence encoding the torsin polypeptide comprising the complete amino acid sequence in SEQ ID NO: 2 or 4;

(b) a nucleotide sequence encoding the torsin polypeptide comprising the complete amino acid sequence encoded by the polynucleotide clone contained in ATCC Deposit Number 98454 or 98455; and

(c) a nucleotide sequence complementary to any of the nucleotide sequences in (a) or (b).

2. The isolated nucleic acid molecule according to claim 1, wherein the molecule comprises the nucleic acid sequence set forth in SEQ ID NO:1 or SEQ ID NO:3.

3. The isolated nucleic acid molecule according to claim 1, wherein the molecule encodes the polypeptide comprising the complete amino acid sequence set forth in SEQ ID NO:2 or 4.

4. The isolated nucleic acid molecule according to claim 1, wherein the nucleotide sequence encoding the torsin polypeptide comprises the complete amino acid sequence encoded by the polynucleotide clone contained in ATCC Deposit Number 98454 or 98455.

5. An isolated nucleic acid molecule consisting of 10 to 50 nucleotides which specifically hybridizes to RNA or DNA encoding torsin protein, wherein said nucleic acid molecule is or is complementary to a nucleotide

sequence consisting of at least 10 consecutive nucleotides from the nucleotide sequence according to claim 1.

6. A method of detecting torsin nucleic acid in a sample comprising:

a) contacting said sample with the nucleic acid molecule according to claim 5, under conditions such that hybridization occurs, and

b) detecting the presence of said molecule bound to torsin nucleic acid.

7. A kit for detecting the presence of torsin nucleic acid in a sample comprising at least one container means having disposed therein the nucleic acid molecule according to claim 5.

8. A recombinant nucleic acid molecule comprising, 5′ to 3′, a promoter effective to initiate transcription in a host cell and the nucleic acid molecule according to claim 1.

9. A recombinant nucleic acid molecule comprising a vector and the nucleic acid molecule according to claim 1.

10. A cell that contains the recombinant nucleic acid molecule according to claim 8.

11. A non-human organism that contains the recombinant nucleic acid molecule according to claim 8.

12. A purified torsin polypeptide having an amino acid sequence at least 90% identical to a sequence selected from the group consisting of:

(a) the amino acid sequence of the torsin polypeptide having the complete amino acid sequence in SEQ ID NO:2 or 4;

(b) the amino acid sequence of the torsin polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 98454 or 98455; and

(c) the amino acid sequence of an epitope-bearing portion of any one of the polypeptides of (a) of (b).

13. A composition comprising the polypeptide of claim 12 and a carrier.

14. An antibody having specific binding affinity to the polypeptide of claim 12.

15. A method of detecting torsin in a sample, comprising:

a) contacting said sample with an antibody according to claim 14, under conditions such that immunocomplexes form, and

b) detecting the presence of said antibody bound to said polypeptide.

16. A diagnostic kit comprising:

a) a first container means containing the antibody according to claim 14 and

b) second container means containing a conjugate comprising a binding partner of said monoclonal antibody and a label.

17. A hybridoma which produces the monoclonal antibody according to claim 14.

18. A bioassay for assessing candidate drugs or ligands of torsin comprising:

a) contacting a candidate drug or ligand with a cell producing the polypeptide according to claim 12; and

b) evaluating the biological activity mediated by said contact.

19. A method of treatment of torsion dystonia in a mammal, comprising administering a therapeutically effective amount of the polypeptide of claim 12 to said mammal.

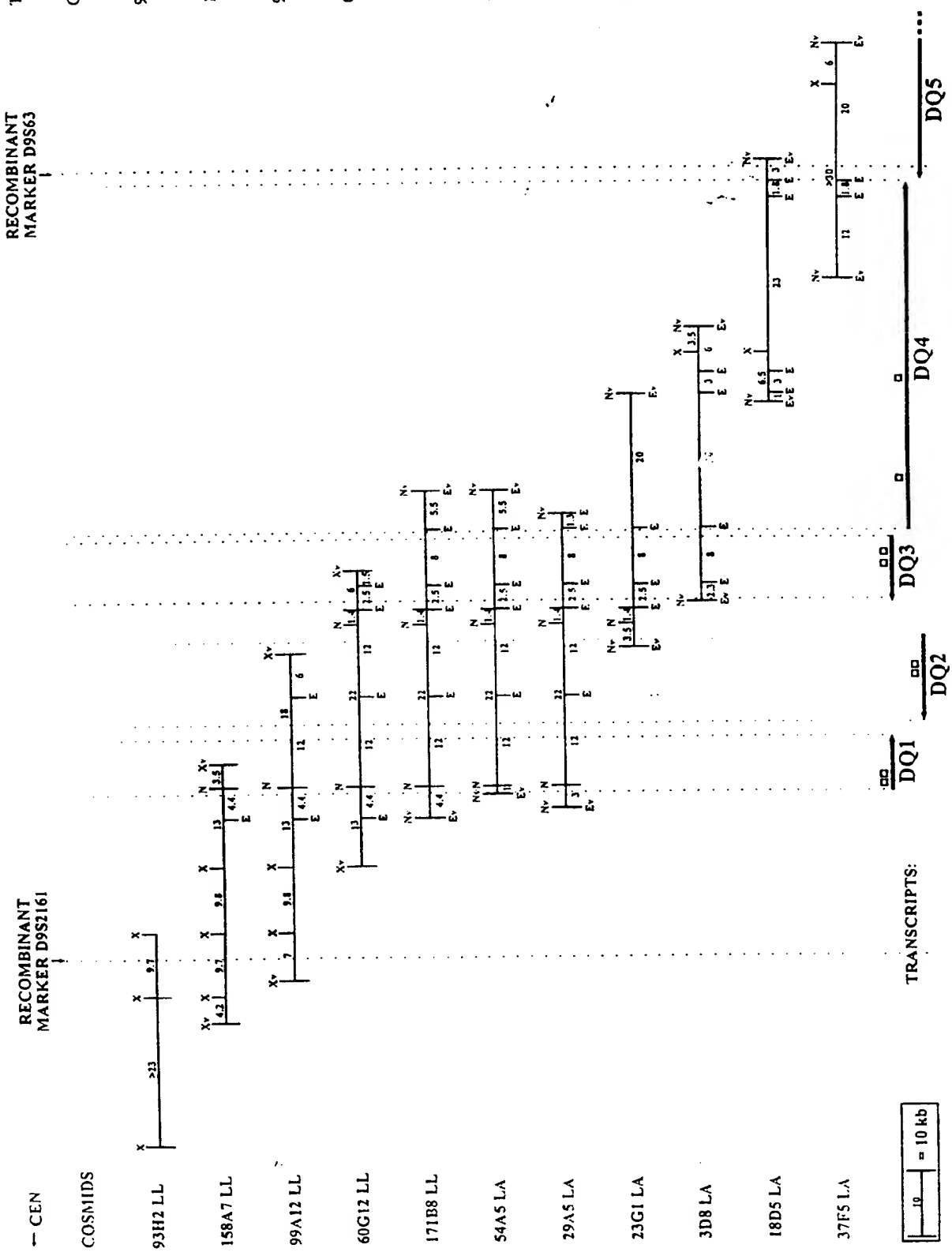
20. A method of diagnosing the presence or predisposition to develop torsion dystonia in a patient, said method comprising:

a) taking a sample from said patient;

b) evaluating the characteristics of torsinA nucleic acid in said sample, wherein said evaluation comprises detecting the GAGGAG region (SEQ ID NO:5 at nucleotide positions 946-951) in said sample; and

c) diagnosing the presence or predisposition to develop torsion dystonia in a patient wherein the absence of a GAG from said GAGGAG region indicates the presence or predisposition to develop torsion dystonia.

FIGURE 1


← CEN
COSMIDS
93H2 LL
158A7 LL
99A12 LL
60G12 LL
171B8 LL
54A5 LA
29A5 LA
23G1 LA
3D8 LA
18D5 LA
37F5 LA
RECOMBINANT MARKER D9S2161
RECOMBINANT MARKER D9S63
TEL →
COSMIDS
93H2 LL
158A7 LL
99A12 LL
60G12 LL
171B8 LL
54A5 LA
29A5 LA
23G1 LA
3D8 LA
18D5 LA
37F5 LA
TRANSCRIPTS:
DQ1
DQ2
DQ3
DQ4
DQ5
= 10 kb

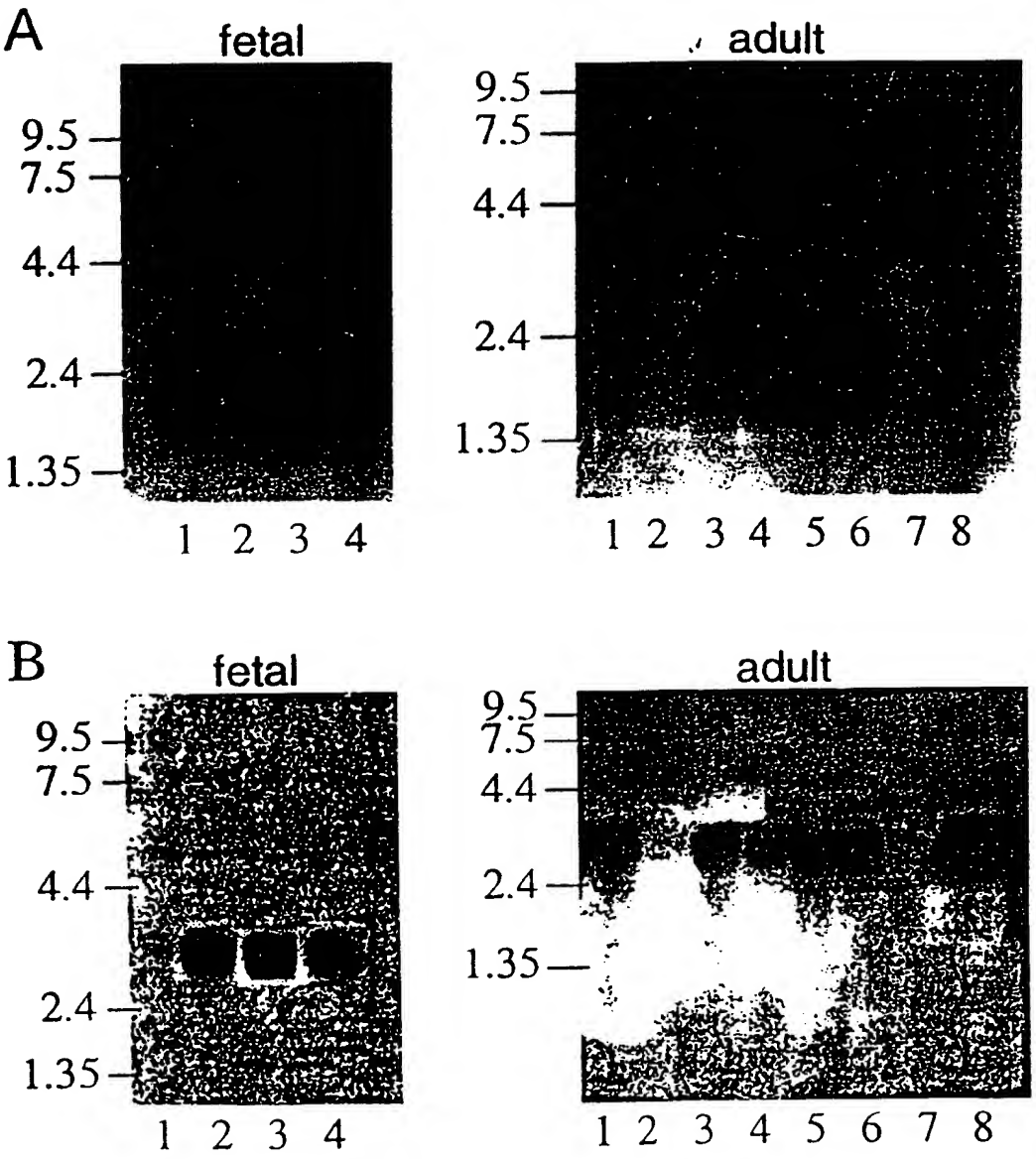

A
fetal
adult
9.5
7.5
4.4
2.4
1.35
1 2 3 4
1 2 3 4 5 6 7 8
B
fetal
adult
9.5
7.5
4.4
2.4
1.35
1 2 3 4
1 2 3 4 5 6 7 8

FIGURE 4

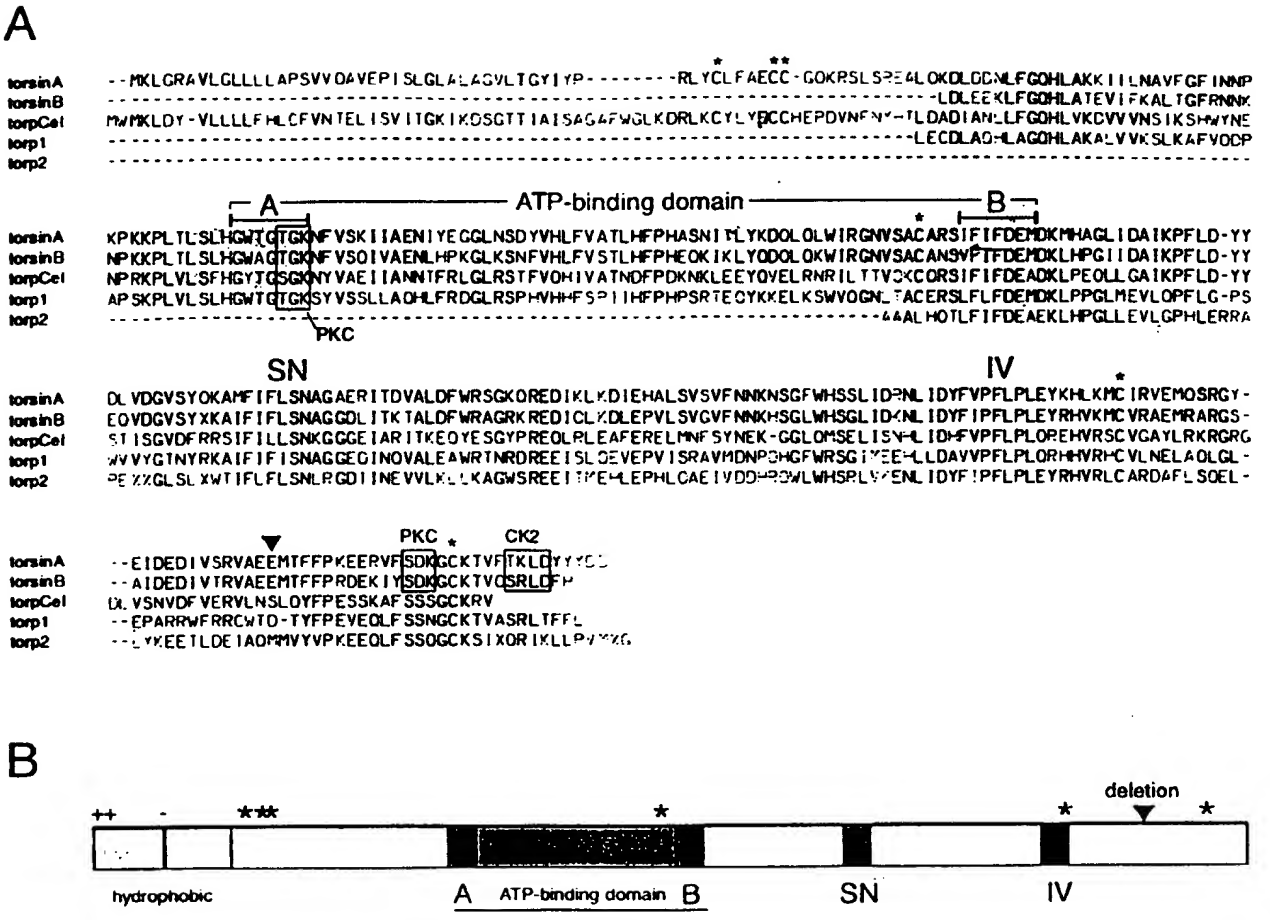
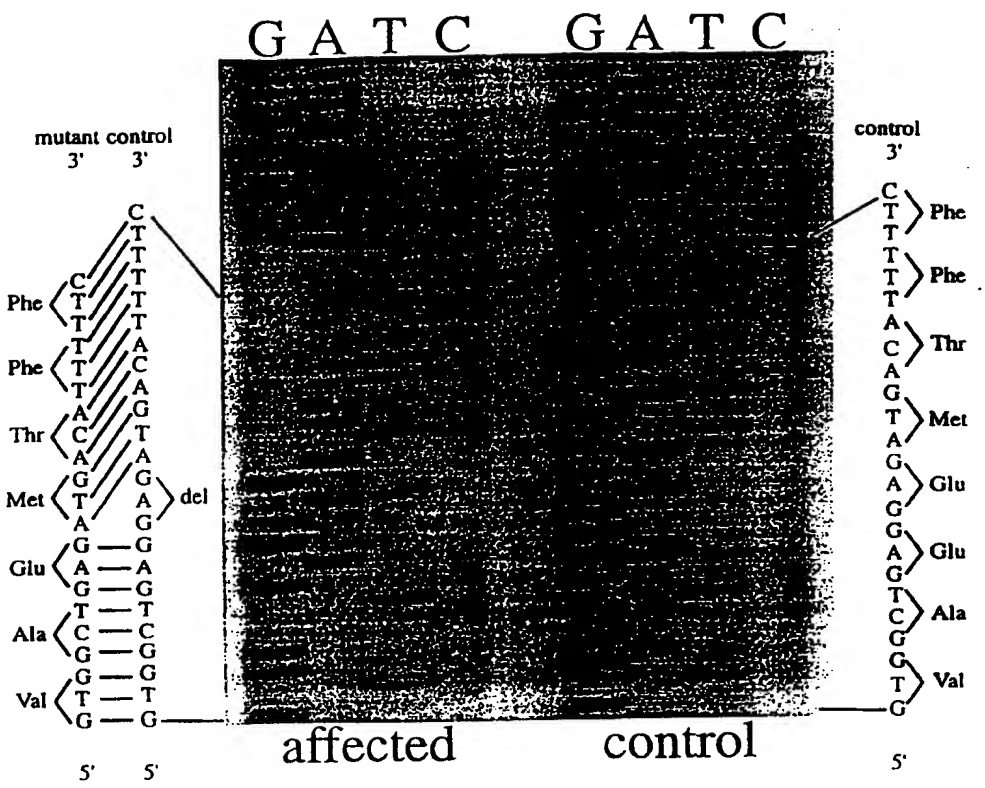

A
torsinA
torsinB
torpCel
torp1
torp2
A
ATP-binding domain
B
PKC
SN
IV
PKC
CK2
B
deletion
hydrophobic
A
ATP-binding domain
B
SN
IV
C
A
B
SKD3
HSP-101
torsinA
torsinB
torpCel
torp1
torp2
SN
IV
SKD3
HSP-101
torsinA
torsinB
torpCel
torp1
torp2

FIGURE 5

A


G A T C G A T C
mutant control
3' 3'
control
3'
Phe
Phe
Thr
Met
del
Glu
Ala
Val
Phe
Phe
Thr
Met
Glu
Glu
Ala
Val
affected
control
5' 5'
5'

B

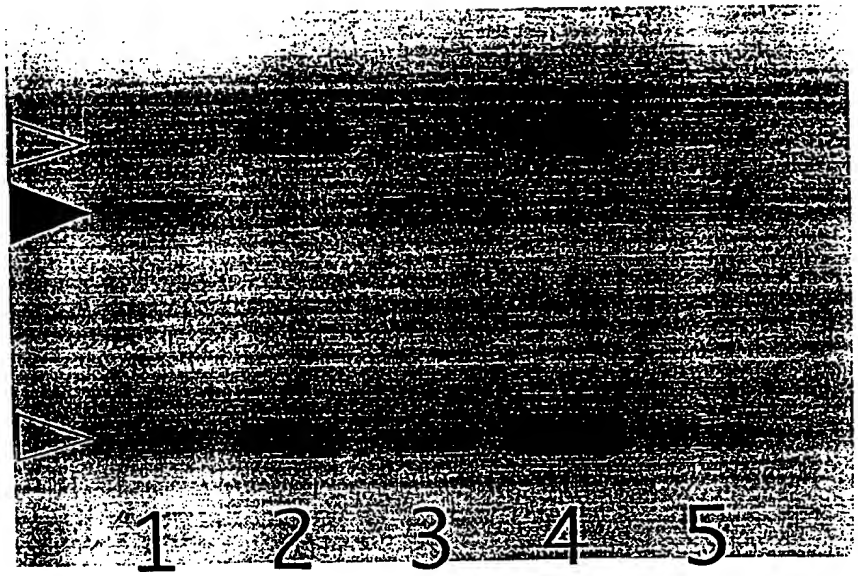
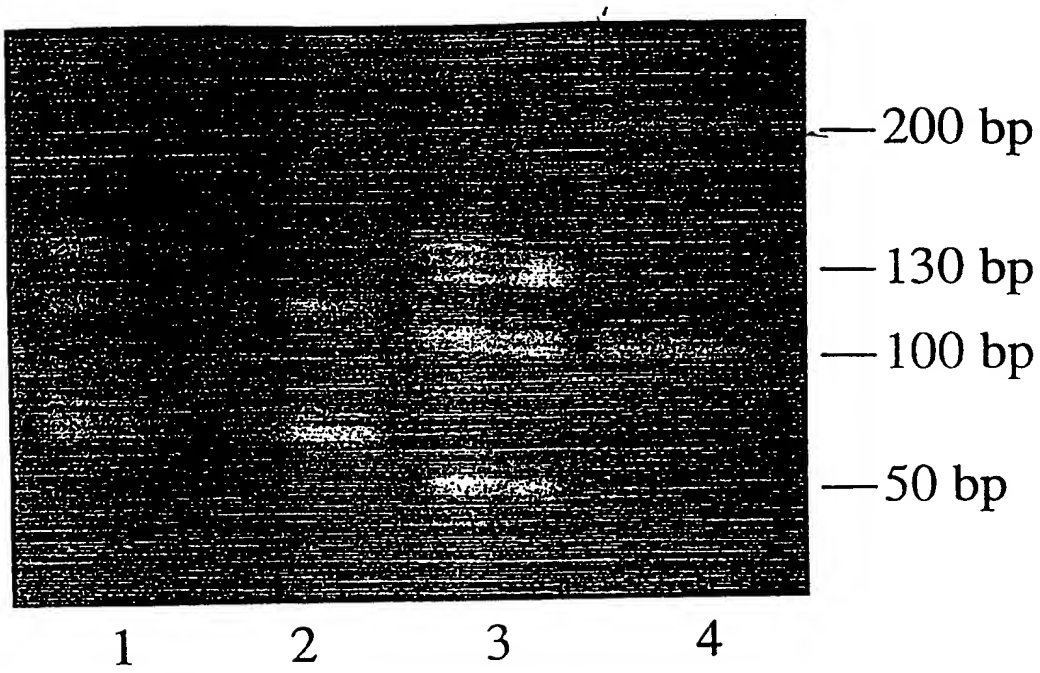
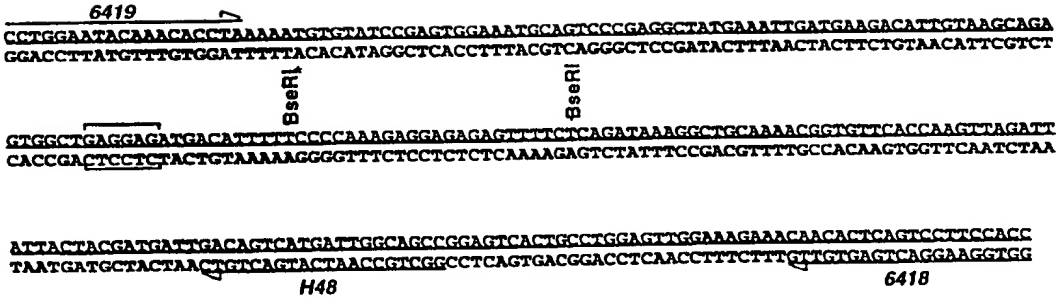

1 2 3 4 5

FIGURE 5

C


200 bp
130 bp
100 bp
50 bp
1
2
3
4

D


6419
CCTGGAATACAAACACCTAAAAATGTGTATCCGAGTGGAAATGCAGTCCCGAGGCTATGAAATTGATGAAGACATTGTAAGCAGA
GGACCTTATGTTTGTGGATTTTTACACATAGGCTCACCTTTACGTCAGGGCTCCGATACTTTAACTACTTCTGTAACATTCGTCT
BseRI
BseRI
GTGGCTGAGGAGATGACATTTTTCCCCAAAGAGGAGAGAGTTTTCTCAGATAAAGGCTGCAAAACGGTGTTCACCAAGTTAGATT
CACCGACTCCTCTACTGTAAAAAGGGGTTTCTCCTCTCTCAAAAGAGTCTATTTCCGACGTTTTGCCACAAGTGGTTCAATCTAA
ATTACTACGATGATTGACAGTCATGATTGGCAGCCGGAGTCACTGCCTGGAGTTGGAAAGAAACAACACTCAGTCCTTCCACC
TAATGATGCTACTAACTGTCAGTACTAACCGTCGGCCTCAGTGACGGACCTCAACCTTTCTTTGTTGTGAGTCAGGAAGGTGG
H48
6418

Applicant's or agent's file reference number 0609.438PC00	International application No. TBA US98/12776

INDICATIONS RELATING TO A DEPOSITED MICROORG
(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page ____16____, line ___14_____.

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☒

Name of depositary institution
American Type Culture Collection

Address of depositary institution *(including postal code and country)*
12301 Parklawn Drive
Rockville, MD 20852
United States of America

Now located at:
10801 University Blvd.
Manassas, VA 20110-2209
United States of America

Date of deposit	Accession Number
June 12, 1997	98454

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

Escherichia coli with cDNA in plasmid H4, DYT1

In respect of those designations in which a European Patent is sought a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample (Rule 28(4) EPC).

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the international Bureau later *(specify the general nature of the indications, e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer Sonya Barnes PCT International Division	Authorized officer

SWEDEN

The applicant hereby requests that, until the application has been laid open to public inspection (by the Swedish Patent Office), or has been finally decided upon by the Swedish Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the International Bureau before the expiration of 16 months from the priority date (preferably on the Form PCT/RO/134 reproduced in annex Z of Volume I of the PCT Applicant's Guide). If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Swedish Patent office or any person approved by the applicant in the individual case.

UNITED KINGDOM

The applicant hereby requests that the furnishing of a sample of a microorganism shall only be made available to an expert. The request to this effect must be filed by the applicant with the International Bureau before the completion of the technical preparations for international publication of the application.

Applicant's or agent's file reference number 0609.438PC0U	International application No. TBA US)98/+12776

INDICATIONS RELATING TO A DEPOSITED MICROORG. . .
(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page ____16_____, line ___14_____.

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution
American Type Culture Collection

Address of depositary institution *(including postal code and country)*
12301 Parklawn Drive
Rockville, MD 20852
United States of America

Now located at:
10801 University Blvd.
Manassas, VA 20110-2209
United States of America

Date of deposit	Accession Number
June 12, 1997	98455

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

Escherichia coli containing human cDNA in plasmid 17

In respect of those designations in which a European Patent is sought a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample (Rule 28(4) EPC).

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the international Bureau later *(specify the general nature of the indications, e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer Sonya Barnes PCT International Division	Authorized officer

CANADA

The applicant hereby requests that, until either a Canadian patent has been issued on the basis of the application or the application has been refused, or is abandoned and no longer subject to reinstatement, or is withdrawn, the furnishing of a sample of deposited biological material referred to in the application only be effected to an independent expert nominated by the Commissioner of Patents.

DENMARK

The applicant hereby requests that, until the application has been laid open to public inspection (by the Danish Patent Office), or has been finally decided upon by the Danish Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Danish Patent Office not later than at the time when the application is made available to the public under Sections 22 and 33(3) of the Danish Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Danish Patent office or any person approved by the applicant in the individual case.

FINLAND

The applicant hereby requests that, until the application has been laid open to public inspection (by the National Board of Patents and Registration), or has been finally decided upon by the National Board of Patents and Registration without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the International Bureau before the expiration of 16 months from the priority date (preferably on the Form PCT/RO/134 reproduced in annex Z of Volume I of the PCT Applicant's Guide). If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the National Board of Patents and Registration or any person approved by the applicant in the individual case.

NETHERLANDS

The applicant hereby requests that until the date of a grant of a Netherlands patent or until the date on which the application is refused or withdrawn or lapsed, the microorganism shall be made available as provided in Rule 31F(1) of the Patent Rules only by the issue of a sample to an expert. The request to this effect must be furnished by the applicant with the Netherlands Industrial Property Office before the date on which the application is made available to the public under Section 22C or Section 25 of the Patents Act of the Kingdom of the Netherlands, whichever of the two dates occurs earlier.